

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:16:14 ; Search time 56.55 Seconds

(without alignments)

799.417 Million cell updates/sec

Title: US-09-724-571-58

Perfect score: 2156

Sequence: 1 ETDEPEEPGRGRGFMVMD.....GPFVILDMEDCGYNIPQTDE 407

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2156	100.0	415	21 AAB07899	Amino acid sequenc
2	2156	100.0	453	21 AAY88438	Modified human asp
3	2156	100.0	453	22 AAEL10642	Human-Asp 2(a) pro
4	2156	100.0	453	22 AAE06872	Human-Asp2(a) delt
5	2156	100.0	453	22 AAU06616	Human-pro-Asp 2(a)
6	2156	100.0	453	22 AAU07215	Human-Asp2(a) del
7	2156	100.0	453	22 AAE02594	Active enzyme port
8	2156	100.0	456	21 AAB07897	Modified human asp
9	2156	100.0	459	21 AAY88439	Human-Asp 2(a) pro
10	2156	100.0	459	22 AAEL10643	Human-Asp2(a) delt
11	2156	100.0	459	22 AAE06873	

12	2156	100.0	459	22 AAU06617	Human-pro-Asp 2(a)
13	2156	100.0	459	22 AAU07216	Human aspartyl pro
14	2156	100.0	459	22 AAE02595	Human-Asp-2(a) del
15	2156	100.0	460	21 AAB07898	Amino acid sequenc
16	2156	100.0	488	22 AAG66572	Human memapsin 2.
17	2156	100.0	488	22 AAB61334	Memapsin 2 protein
18	2156	100.0	501	21 AAY94767	Human beta-secreta
19	2156	100.0	501	22 AAB07896	Amino acid sequenc
20	2156	100.0	503	22 AAB66573	Human pro-memapsin
21	2156	100.0	503	22 AAB61335	T7 promoter and ve
22	2151	99.8	433	21 AAY88433	Human-pro-Asp-2(a)
23	2151	99.8	433	22 AAEL10640	Human-pro-Asp 2(a)
24	2151	99.8	433	22 AAE06870	Human-pro-Asp2(a)
25	2151	99.8	433	22 AAU06614	Human-pro-Asp 2(a)
26	2151	99.8	433	22 AAU07213	T7-human aspartyl
27	2151	99.8	433	22 AAE02592	Human-pro-Asp-2(a)
28	2151	99.8	446	21 AAY88431	Human-pro-Asp-2(a)
29	2151	99.8	446	22 AAEL10638	T7-caspase-human-p
30	2151	99.8	446	22 AAE06868	T7-human-pro-Asp 2
31	2151	99.8	446	22 AAU06612	Human T7-human-pro
32	2151	99.8	446	22 AAU07211	T7-human aspartyl
33	2151	99.8	446	22 AAE02590	T7-human-pro-Asp-2
34	2151	99.8	459	21 AAY88432	T7-caspase-human-p
35	2151	99.8	459	22 AAEL10639	T7-Caspase-human-p
36	2151	99.8	459	22 AAE06869	T7-Caspase-Human-p
37	2151	99.8	459	22 AAU06613	Human T7-Caspase-H
38	2151	99.8	459	22 AAU07212	T7-Caspase-human-a
39	2151	99.8	459	22 AAE02591	T7-Caspase-human-p
40	2151	99.8	501	21 AAY88425	Human aspartyl pro
41	2151	99.8	501	22 AAE10629	Human aspartyl pro
42	2151	99.8	501	22 AAE06859	Human aspartyl pro
43	2151	99.8	501	22 AAU06503	Human aspartyl pro
44	2151	99.8	501	22 AAU07202	Human aspartyl pro
45	2151	99.8	501	22 AAE02581	Human aspartyl pro

ALIGNMENTS

RESULT 1  
AAB07899  
ID AAB07899 standard; Protein; 415 AA.  
XX AC AAB07899;  
XX DT 14-NOV-2000 (first entry)  
XX DE Amino acid sequence of a human beta-secretase enzyme fragment.  
XX KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
XX KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
XX KW inhibitor.  
XX OS Homo sapiens.  
XX PN WO200047618-A2.  
XX PD 17-AUG-2000.  
XX PF 10-FEB-2000; 2000WO-US03819.  
XX PR 10-FEB-1999; 99US-0119571.  
XX PR 15-JUN-1999; 99US-0119172.  
XX PA (ELAN-) ELAN PHARM INC.  
XX PI Anderson JP, Basl G, Doane MT, Frigon N, John V, Power M;  
XX PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
XX WPI; 2000-533011/48.  
XX DR Purified beta-secretase protein used in assays to discover inhibitors  
XX PT which can be used for the treatment of amyloidogenic diseases e.g.

```
PT Alzheimer's disease -
XX
XX Claim 10; Fig 3B; 121pp; English.
XX
CC The specification describes a beta-secretase enzyme. The enzyme cleaves
CC beta-amyloid precursor protein to produce beta-amyloid peptide. This
CC enzyme is therefore implicated in the production of amyloid plaque
CC components which accumulate in the brains of individuals afflicted with
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's
CC disease-like pathology to test if they maintain or improve cognitive
CC ability or reduce the plaque burden. The compounds are used for the
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The
CC present sequence represents a human beta-secretase enzyme fragment.
XX
XX Sequence 415 AA;
SQ
Query Match 100.0%; Score 2156; DB 21; Length 415;
Best Local Similarity 100.0%; Pred. No. 2.1e-210;
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ETDEPEEPEGRGSGFVEMVDNLRGKSGQGYVEMTVGSPPTNLILVDTGSSNFVAGAAP 60
Db 1 etdeepeepegrgsgfvmvndlrgksggyvemtvgspptnlilvdtgssnfavgaap 60
QY 61 HPFLHRYQRLSTYRDLRGVVPYTOGKWEGLCTDLVSIHPGPNVTVRANIAAITE 120
Db 61 hpflhryyqrlslyrdlrgvvyptqgkwegeigtdlvsiphgpnvtvranaiaite 120
QY 121 SDKFFINGSNWEGILGLAYAEIARPDLSLEPFFDLSLVKQTHVPLNLSLQLCGAGFPLNQS 180
Db 121 sdkffingsnwegilglayaeiarpddslepffdsllvkqthvplnlsfqlcgaagfplnqs 180
QY 181 EVLASVGSMIIGIDHSILYTGSLWYTPIRREWYVEVLIIVRVEINGQDLKMDCKEYNVDK 240
Db 181 evlasvgsgmiigidhsilytgslwytpirrewyvevliivrveingqdlkmdckeynydk 240
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIKAASSTKFPDGLFWLGQWQAGTTPWNIFPVIS 300
Db 241 sivdsgttnlrpkkvfeaaavksikaasstekfpdgfwlgeqlvcwqagttwnifpvis 300
QY 301 LYLMEVNTQSFRTIILPQQYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYVV 360
Db 301 lylmgevntqsfritiilpqyylrpvedvatsddcykfaisqsstgtvmgavimegyfvv 360
QY 361 FDRARKRIGFAVSACHVHDEPRTAAVEGPFVTLDMDCGYNIPQDE 407
Db 361 fdrarkrigfavsachvhdeirtaavegpfvtldmedcgyynipqde 407
RESULT 2
AAE106438
ID AAE106438 standard; Protein; 453 AA.
XX
XX AAE106438;
XX
XX 03-AUG-2000 (first entry)
XX
XX Modified human aspartyl protease 2 (Asp2) amino acid sequence.
XX
XX Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
XX Alzheimer's disease; beta secretase site.
XX
XX Homo sapiens.
XX
XX WO200017369-A2.
XX
XX 30-MAR-2000.
XX
XX 23-SEP-1999; 99WO-US20881.
XX
XX 24-SEP-1998; 98US-0101594.
XX
```

```
XX
XX (PHAA ) PHARMACIA & UPJOHN CO.
XX
XX Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
XX
XX WPI; 2000-303209/26.
XX N-PSDB; AAA15688.
XX
XX New enzyme designated human aspartase useful in research into
XX Alzheimer's Disease is capable of cleaving amyloid protein precursor at
XX the beta secretase site to produce amyloid beta peptide -
XX
XX Example 10; Page 169-172; 183pp; English.
XX
XX This sequence represents a modified human aspartyl protease 2 (Asp2)
XX amino acid sequence. Asp2 encoded by this sequence has the C-terminal
XX transmembrane domain deleted. The invention relates to a protease
XX (e.g. Asp2) capable of cleaving the beta secretase site of amyloid
XX precursor protein (APP). The protease contains a sequence encoding the
XX amino acid sequence DTG and a sequence encoding DSG or DTG separated by
XX 100-300 amino acids. When mutated the APP gene causes an autosomal
XX dominant form of Alzheimer's disease. APP localises to the cell surface
XX membrane and have a single C-terminal transmembrane domain. Proteolytic
XX processing of APP produces the amyloid beta protein, which is possibly
XX very important in Alzheimer's disease. The invention includes a
XX nucleotide sequence encoding the protease, a vector containing the
XX nucleotide sequence, and a cell line comprising the vector. Methods for
XX screening for inhibitors of beta secretase activity are also given in the
XX invention. The human aspartase protein and nucleotide sequences and the
XX methods for identifying inhibitors of the protease, are useful in the
XX treatment of and research in to Alzheimer's disease.
XX
XX Sequence 453 AA;
SQ
Query Match 100.0%; Score 2156; DB 21; Length 453;
Best Local Similarity 100.0%; Pred. No. 2.4e-210;
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ETDEPEEPEGRGSGFVEMVDNLRGKSGQGYVEMTVGSPPTNLILVDTGSSNFVAGAAP 60
Db 46 etdeepeepegrgsgfvmvndlrgksggyvemtvgspptnlilvdtgssnfavgaap 105
QY 61 HPFLHRYQRLSTYRDLRGVVPYTOGKWEGLCTDLVSIHPGPNVTVRANIAAITE 120
Db 106 hpflhryyqrlslyrdlrgvvyptqgkwegeigtdlvsiphgpnvtvranaiaite 165
QY 121 SDKFFINGSNWEGILGLAYAEIARPDLSLEPFFDLSLVKQTHVPLNLSLQLCGAGFPLNQS 180
Db 166 sdkffingsnwegilglayaeiarpddslepffdsllvkqthvplnlsfqlcgaagfplnqs 225
QY 181 EVLASVGSMIIGIDHSILYTGSLWYTPIRREWYVEVLIIVRVEINGQDLKMDCKEYNVDK 240
Db 226 evlasvgsgmiigidhsilytgslwytpirrewyvevliivrveingqdlkmdckeynydk 285
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIKAASSTKFPDGLFWLGQWQAGTTPWNIFPVIS 300
Db 286 sivdsgttnlrpkkvfeaaavksikaasstekfpdgfwlgeqlvcwqagttwnifpvis 345
QY 301 LYLMEVNTQSFRTIILPQQYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYVV 360
Db 346 lylmgevntqsfritiilpqyylrpvedvatsddcykfaisqsstgtvmgavimegyfvv 405
QY 361 FDRARKRIGFAVSACHVHDEPRTAAVEGPFVTLDMDCGYNIPQDE 407
Db 406 fdrarkrigfavsachvhdeirtaavegpfvtldmedcgyynipqde 452
RESULT 3
AAE10642
ID AAE10642 standard; Protein; 453 AA.
XX
XX AAE10642;
XX
```



CC Human aspartyl proteases can act as beta-secretase proteases useful for  
CC treating Alzheimer's disease. App isoforms are useful for identifying  
CC modulators of amyloid-beta peptide production, for use in designing  
CC therapeutics for the treatment and prevention of Alzheimer's disease,  
CC dementia, formation of amyloid plaques, neurofibrillary tangles, gliosis  
CC and neuronal loss. APP isoforms are also used in methods for identifying  
CC inhibitors and modulators of human Asp2 activity. The invention relates  
CC to a method for identifying agents that modulate the activity of human  
CC aspartyl protease Asp2. Amyloid-beta peptides obtained from APP are used  
CC as a means to screen in cellular assays for the inhibitors of beta- and  
CC gamma- secretase. Hu-Asp DNA fragments are useful as probes or primers in  
CC polymerase chain reactions (PCR). The probes are useful for detecting  
CC Hu-Asp nucleic acids in in vitro assays and in Northern and Southern  
CC blots. The present sequence is Human aspartyl protease 2a (Hu-Asp2a)  
CC deltaTM protein which is obtained by the deletion of transmembrane domain  
CC at the C-terminal end of Hu-Asp2a. Human Asp2a has beta-secretase  
CC activity.

XX SQ Sequence 453 AA;

Query Match 100.0%; Score 2156; DB 22; Length 453;  
Best Local Similarity 100.0%; Pred. No. 2.4e-210;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEPGRGRGSEFVEMVDNLKSGGYYVEMTVGSPQTLLNILDVTGSSNFVGAAP 60  
Db 46 etdeepepgrgrgsfvmvndlrgksqgyvemtvgspqtnilvdtgssnfavgaap 105  
QY 61 HPFLHRYQROLSTFYRLDRKGVVPTQGWEGELGDLVSIPHGNVTVRANIAAITE 120  
Db 106 hpflhryyqrlstfyrdlrgkgyvpytqgkwegeldlvsiphgpnvtvranaiaite 165  
QY 121 SKDFIFNGSNWEGILGAYABIAIPDDSLPEFFDSLVKQTHVPMFLSLQLCGAGFFLNQS 180  
Db 166 sdkffingsnwegilgayaaiarpddslpeffdslvkqthvpmflslqcgagfflnqs 225  
QY 181 EVLASVGSMLIGGIDHSLYGSLWYTPIRREWYVEVLIIVRVEINGDQDKMCKEYNYDK 240  
Db 226 evlasvgsmliggidhslygslwypirrewyvevliivrveingdqdkmckeynydk 285  
QY 241 SIVDSGTTNLRPKVFAAVKSIKAASSTKFPDGFGLGQLVCWQAGTTPWNIFFVIS 300  
Db 286 sivdsgettlnlrpkvfaavksikaasstekfpdgfglwglqvcwqagttppwniffvis 345  
QY 301 LYLMGEVTVNSFRITLPQQYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 360  
Db 346 lyimgvntnqsfritilpqqlrprvedvatsqddcykfaissstgtvmgavimegyfvyv 405  
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTD 407  
Db 406 fdrarkrigfavsachvhdefrtaavegpfvltlmedcgynipqtd 452

RESULT 5  
AAU06616  
ID AAU06616 standard; Protein: 453 AA.

XX AC AAU06616;

XX DT 24-OCT-2001 (first entry)

XX DE Human-pro-Asp 2(a) delta TM.

KW Human; Aspartyl protease; beta-secretase; neurotropic; ASP2;

KW neuroprotective; amyloid protein precursor; App; Alzheimer's disease;

KW amyloid-beta; Abeta; Human-pro-Asp 2(a) delta TM; mutant; muten.

XX OS Homo sapiens.

OS Synthetic.

XX FH Key Location/Qualifiers

FT Peptide 1..21

FT Protein /label= Signal\_peptide  
FT 22..453 /label= Mature\_Human\_pro\_Asp\_2(a)\_delta\_TM  
FT Misc-difference 214 /note= "Encoded by CAC"  
XX WO200149098-A2.  
XX 12-JUL-2001.  
XX 09-MAY-2001; 2001WO-IB00798.  
XX 09-MAY-2001; 2001WO-IB00798.  
XX (BIEN/) BIENKOWSKI M J.  
XX (GURN/) GURNEY M E. J.  
XX (HEIN/) HEINRIKSON R L.  
XX (PARO/) PARODI L A.  
XX (YANR/) YAN R.  
XX Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;  
XX WPI; 2001-502549/55.  
XX N-PSDB; AAS11530.  
XX Novel purified polypeptide comprising fragment of mammalian aspartyl  
XX protease 2, lacking Asp2 transmembrane domain and retaining beta  
XX secretase activity of Asp2 useful for identifying inhibitors of Asp2  
XX activity -  
XX Claim 149; Page 160; 185pp; English.  
XX The invention relates to a purified polypeptide comprising a fragment of  
XX mammalian aspartyl protease (Asp)2 protein which lacks the Asp2  
XX transmembrane domain and the Asp2 protein, and where the polypeptide and  
XX the fragment retain the beta-secretase activity of the mammalian Asp2  
XX protein. The invention also details polynucleotides for the Asp  
XX proteins and vectors expressing them, and a polypeptide (isoform of  
XX amyloid protein precursor (APP)) comprising the amino acid sequence of an  
XX APP or its fragment containing an APP cleavage site recognizable by a  
XX mammalian beta-secretase, and further comprising two lysine residues at  
XX the carboxyl terminus of the amino acid sequence of the mammalian APP or  
XX APP fragment. Also included in the invention are methods of identifying  
XX modulators or inhibitors of Asp2. Modulators and inhibitors of Asp2 are  
XX useful for treating Alzheimer's disease. APP is useful in methods for  
XX identifying inhibitors or modulators of human Asp2 activity and  
XX amyloid-beta (Abeta) peptide production. APP is also useful in designing  
XX therapeutics for the treatment or prevention of Alzheimer's disease.  
XX APP comprising the APP-Sw-beta-secretase peptide sequence (NLDA), which  
XX is associated with increased levels of Abeta processing is useful in  
XX assays relating the Alzheimer's research. The expression vector is useful  
XX for recombinantly expressing APP. Nucleic acids that hybridize to  
XX Asp oligonucleotides are useful as probes or primers. The probes are  
XX useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
XX Northern and Southern blots. The present sequence is Human-pro-  
XX Asp 2(a) delta TM protein, which lacks the C-terminal transmembrane  
XX domain.

XX SQ Sequence 453 AA;

Query Match 100.0%; Score 2156; DB 22; Length 453;  
Best Local Similarity 100.0%; Pred. No. 2.4e-210;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEPGRGRGSEFVEMVDNLKSGGYYVEMTVGSPQTLLNILDVTGSSNFVGAAP 60  
Db 46 etdeepepgrgrgsfvmvndlrgksqgyvemtvgspqtnilvdtgssnfavgaap 105

QY 61 HPFLHRYQROLSTFYRLDRKGVVPTQGWEGELGDLVSIPHGNVTVRANIAAITE 120  
Db 106 hpflhryyqrlstfyrdlrgkgyvpytqgkwegeldlvsiphgpnvtvranaiaite 165

QY 121 SDKFFINGSNWEGILGAYAEIARPDSDLPPFFDSLVKQTHVPNLSLQLCGAGFPNQS 180  
 Db |||||||  
 QY 166 sdkffingsnwegilglayaeiarpdslpffdsivkqchvnlslqlcgagfplngs 225  
 Db |||||||  
 QY 181 EVLASVGSMLIGIDHSLYTGSLWTPIRREWYEVIIIVRVEINGQDLKMDCKEYNDK 240  
 Db |||||||  
 QY 226 evlasvgssmligdhsltytgslywtptirrewyevliivveingqdlkmdckeyndk 285  
 QY 241 SIYDSGTTNLRPKKVFEEAAVKSIIKAASSTKEKPPDGFNLGEOQLVCHQAGTTPWNIPFVIS 300  
 Db |||||||  
 QY 286 slvdsqgttnlrpkkvfeavksikaasctekfpdgfwlgeqlvcwqagttppwnifpvis 345  
 QY 301 LYLMEVTNQSPRITILPQQYLRPVEDVATSDQCYKFAISQSSGTVMGAVIMEGFYV 360  
 Db |||||||  
 QY 346 lylmgevtngsfritilpqgylrpvedvatsqdcykfaissgstgtvmgavimegfvyv 405  
 QY 361 FDRARKRIGFAYSAHVHDEFTAAVEGPFVTLDMEDCGYNIPQDTE 407  
 Db |||||||  
 QY 406 fdrarkrigfavsachvhdeftaavegpfvltlmedcgynipqtde 452

RESULT 6  
 AAU07215  
 ID AAU07215 standard; Protein; 453 AA.  
 XX  
 AC AAU07215;  
 XX  
 DT 24-OCT-2001 (first entry)  
 XX  
 DE Human aspartyl protease 2a deltaTM (HuAsp-2adeltaTM).  
 XX  
 KW Human; aspartyl protease 1; Asp-1; nootropic; neuroprotective;  
 KW aspartyl protease 2; Asp2; amyloid protein precursor; APP;  
 KW beta-secretase; Alzheimer's disease; HuAsp-2adeltaTM.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT Peptide  
 FT 1..21 Location/Qualifiers  
 FT /note= "Signal peptide"  
 FT 22..453  
 FT /note= "Mature human aspartyl protease 2a deltaTM"  
 PN WO200149097-A2.  
 XX  
 XX 12-JUL-2001.  
 XX  
 XX 09-MAY-2001; 2001WO-IB00797.  
 XX  
 XX 09-MAY-2001; 2001WO-IB00797.  
 PR  
 XX (BIEN/) BIENKOWSKI M J.  
 PA (GURN/) GURNEY M E.  
 PA (HEIN/) HEINRIKSON R L.  
 PA (PARO/) PARODI L A.  
 PA (YANR/) YAN R.  
 XX  
 PI Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;  
 XX  
 DR WPI: 2001-502548/55.  
 DR N-PSDB; AAS11715.  
 XX  
 XX Novel purified polypeptide comprising fragment of mammalian aspartyl  
 PT protease 2, lacking Asp2 transmembrane domain and retaining beta  
 PT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
 PT activity -  
 XX  
 XX Claim 149; Fig 11; 185pp; English.  
 PS  
 XX The invention relates to a novel purified polypeptide comprising a  
 CC fragment of mammalian aspartyl protease 2 (Asp2) protein which lacks the  
 CC Asp2 transmembrane domain and the Asp2 protein, and where the polypeptide  
 CC and the fragment retain the beta-secretase activity of the mammalian Asp2

protein. Also included is an isoform of amyloid protein precursor (APP) comprising the amino acid sequence of a APP or its fragment containing an APP cleavage site recognizable by a mammalian beta-secretase, and further comprising two lysine residues at the carboxyl terminus of the amino acid sequence of the mammalian APP or APP fragment. The polypeptides are used for assaying for modulators of beta-secretase activity; identifying agents that inhibit the APP processing activity of human Asp2 aspartyl protease (Hu-Asp2); identifying agents that modulate the activity of Asp2; and for reducing cellular production of amyloid beta (Abeta) from APP. Agents identified by the above methods are useful for treating Alzheimer's disease; and for identifying modulators of amyloid-beta (Abeta) peptide production, for use in designing therapeutics for the treatment or prevention of Alzheimer's disease. Probes and primers derived from Asp nucleic acid sequences are useful for detecting Hu-Asp nucleic acids in in vitro assays and Northern and Southern blots. The present sequence represents the amino acid sequence of human Asp-2a delta TM construct which lacks the transmembrane domain. This construct was used for bacterial expression and purification of human Asp2a.

XX  
 SQ Sequence 453 AA;

Query Match 100.0%; Score 2156; DB 22; Length 453;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-210;  
 Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDPEEPGRGRGSFVEMVDNLRGKSGQGYVEMTVGSPPTLNILVDVGGSSNFAVGAAP 60  
 Db |||||||  
 QY 46 etdeepeepgrgrgsfvmvndnlrgksggyvemtvgspptlnilvdtgssnfavgaap 105  
 Db |||||||  
 QY 61 HPFLHRYQRQLSSTYRDLRKGVVYPYTGCKWEGELGTLVSIPIHGPNTVVRANIAAITE 120  
 Db |||||||  
 QY 106 hpflhryyqrqlsstyrdlrkgyvpytgckwegelgtlvsiphgpnvtvranaiaite 165  
 QY 121 SDKFFINGSNWEGILGAYAEIARPDSDLPPFFDSLVKQTHVPNLSLQLCGAGFPNQS 180  
 Db |||||||  
 QY 166 sdkffingsnwegilglayaeiarpdslpffdsilvkqthvnlslqlcgagfplngs 225  
 QY 181 EVLASVGSMLIGIDHSLYTGSLWTPIRREWYEVIIIVRVEINGQDLKMDCKEYNDK 240  
 Db |||||||  
 QY 226 evlasvgssmligdhsltytgslywtptirrewyevliivveingqdlkmdckeyndk 285  
 QY 241 SIYDSGTTNLRPKKVFEEAAVKSIIKAASSTKEKPPDGFNLGEOQLVCHQAGTTPWNIPFVIS 300  
 Db |||||||  
 QY 286 slvdsqgttnlrpkkvfeavksikaasctekfpdgfwlgeqlvcwqagttppwnifpvis 345  
 QY 301 LYLMEVTNQSPRITILPQQYLRPVEDVATSDQCYKFAISQSSGTVMGAVIMEGFYV 360  
 Db |||||||  
 QY 346 lylmgevtngsfritilpqgylrpvedvatsqdcykfaissgstgtvmgavimegfvyv 405  
 QY 361 FDRARKRIGFAYSAHVHDEFTAAVEGPFVTLDMEDCGYNIPQDTE 407  
 Db |||||||  
 QY 406 fdrarkrigfavsachvhdeftaavegpfvltlmedcgynipqtde 452

RESULT 7  
 AAU02594  
 ID AAU02594 standard; Protein; 453 AA.  
 XX  
 AC AAU02594;  
 XX  
 DT 10-AUG-2001 (first entry)  
 XX  
 DE Human-Asp-2(a) delta TM protein.  
 XX  
 KW Human; alpha-secretase; amyloid precursor protein; APP; therapy;  
 KW Alzheimer's disease; antialzheimer's; aspartyl protease 2a; Asp2a;  
 KW beta-secretase; Asp-2a delta TM.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.

Key	Location/Qualifiers	
FT	Misc-difference 214	
FT	/note= "Encoded by CAC"	
FT		
PN	WO200123533-A2.	
PN		
PD	05-APR-2001.	
XX	22-SEP-2000; 2000WO-US26080.	
XX	23-SEP-1999; 99US-0155493.	
XX	23-SEP-1999; 99WO-US20881.	
PR	13-OCT-1999; 99US-0416901.	
PR	06-DEC-1999; 99US-0169232.	
XX	(PHAA ) PHARMACIA & UPTJOHN CO.	
XX		
XX	Gurney M, Bienkowski MJ;	
XX	WPI: 2001-290516/30.	
XX	N-PSDB; AAD06752.	
DR		
XX	Enzymes that cleave the alpha-secretase site of the amyloid precursor protein, useful for the treatment of Alzheimer's disease -	
XX	Example 10; Fig 11; 189pp; English.	
XX	The present invention relates to enzymes for cleaving the alpha-secretase site of the amyloid precursor protein (APP) and methods of identifying those enzymes. The methods may be used to identify enzymes that may be used to cleave the alpha-secretase cleavage site of the APP protein. The enzymes may be used to treat or modulate the progress of Alzheimer's disease. The present sequence is human Aspartyl protease 2a (Asp-2a) deltatM protein which is obtained by deleting its transmembrane domain. This sequence has beta-secretase protease activity.	
XX	Sequence 453 AA;	
SQ		
Query Match	100.0%; Score 2156; DB 22; Length 453;	
Best Local Similarity	100.0%; Pred. No. 2.4e-210;	
Matches 407; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 ETDEPEPPGRGSRFVEMVDNLRGKSGQGYVEMTVGSPPTLNILVDTGSSNFVAGAAP 60	
Db	46 etdeepepggrgsfvmvndnlrgksggyvemtvgspptlnilvdtgssnfavgaap 105	
QY	61 HPFLHRYQRLSSTYRDLRGVVPYTOGKWEGLGTLVSPHGPNTVVRANIAAITE 120	
Db	106 hpflhryyqrlsstyrdlrgvvyptqgkwegeigtldvslphgpnvtvranaiaite 165	
QY	121 SDKFFINGSNWEGILGLAYAEIARPDSDLSPFFDLSLVKQTHVPHNLSLQICGAGFPLNOS 180	
Db	166 sdkffingsnwegilglayaeiarpdslslepfdfsllvkqthvphnlsqicgagfplns 225	
QY	181 EVLASVGSMTIGGDHSLYTGSLWYTPIRREWYVEIIVRVEINGDLKMDCKEYNYDK 240	
Db	226 evlasvgsmtiiggdhsltygslwytpirrewyveilvrveingdltkmdckeynydk 285	
QY	241 SIVDSGTTNLRPKVFEAAVKSIIKAASSTKEKFDGFWLGEQLVCWQAGTTPNNIFPVIS 300	
Db	286 sivdsgettlnlrpkvfeavksikaasstekfdgfwlgeqlvcwqagttppnnifpvis 345	
QY	301 LYLMEGVTNQSFRITILFOQYLRPVEDVATSQDDCYKFAISOSSTGVMGAVIMEGFYVV 360	
Db	346 lylmgevtnqsfritilpqyrlrpyedvatseqddcykfaissqstgtvmgavimegyvv 405	
QY	361 FDRARIRIGFAYSACHVHDEFRTAAVEGPFVITLDMEDCGYNIPTDE 407	
Db	406 fdraririgfavsachvhdefrtaaavegpfvltlmedcgynipqtde 452	
RESULT	8	

Db 241 sivdsgrtlnrlpkkvfeavksikaasstekfpgfwlgeqlvcwgaattpwlnifpvis 300  
QY 301 LVLGSEVTNQSFRITILPQOYLRPVEDVATSDDCYKFAISOSSCTGTVMGAVIMEGFYV 360  
Db 301 lylmgevtnqsfritilpqgylrpvedvatsqddcykfaissstgtvmgavimegfyyv 360  
QY 361 FDRARRKRGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDE 407  
Db 361 fdrarkrigrfavsachvhdefrtaavegpfvtilmedecgynipqtde 407  
RESULT 9  
AAE10643 9  
ID AAY88439 standard; Protein; 459 AA.  
AC AAY88439;  
XX  
DT 03-AUG-2000 (first entry)  
XX  
DE Modified human aspartyl protease 2 (Asp2) amino acid sequence.  
XX  
KW Aspartyl protease; aspartase; amyloid precursor protein; App; Asp 2;  
KW Alzheimer's disease; beta secretase site.  
XX  
OS Homo sapiens.  
XX  
PN WO200017369-A2.  
XX  
PD 30-MAR-2000.  
XX  
PF 23-SEP-1999; 99WO-US20881.  
XX  
PR 24-SEP-1998; 98US-0101594.  
XX  
PA (PHAA ) PHARMACIA & UPJOHN CO.  
XX  
PI Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;  
XX  
DR WPI; 2000-303209/26.  
DR N-PSDB; AAA15689.  
XX  
PT New enzyme designated human aspartase useful in research into  
PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at  
PT the beta secretase site to produce amyloid beta peptide -  
XX  
PS Example 10; Page 173-176; 183pp; English.  
XX  
CC This sequence represents a modified human aspartyl protease 2 (Asp2)  
CC amino acid sequence. Asp2 encoded by this sequence has the C-terminal  
CC transmembrane domain deleted. The invention relates to a protease  
CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid  
CC precursor protein (APP). The protease contains a sequence encoding the  
CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by  
CC 100-300 amino acids. When mutated the APP gene causes an autosomal  
CC dominant form of Alzheimer's disease. APP localises to the cell surface  
CC membrane and have a single C-terminal transmembrane domain. Proteolytic  
CC processing of APP produces the amyloid beta protein, which is possibly  
CC very important in Alzheimer's disease. The invention includes a  
CC nucleotide sequence encoding the protease, a vector containing the  
CC nucleotide sequence, and a cell line comprising the vector. Methods for  
CC screening for inhibitors of beta secretase activity are also given in the  
CC invention. The human aspartase protein and nucleotide sequences and the  
CC methods for identifying inhibitors of the protease, are useful in the  
CC treatment of and research in to Alzheimer's disease.  
XX  
SQ Sequence 459 AA;  
Query Match 100.0%; Score 2156; DB 21; Length 459;  
Best Local Similarity 100.0%; Pred. No. 2.5e-210;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEPEGRGSGFVEMVDNLCKSGOGYXVEMTVGSPQTNLILVDTGSSNFAVGAAP 60  
Db 46 etdepeepgrgrsgfvmvndnlrgksggyvemtvgspqtnlilvdtgsnfavgaap 105  
QY 61 HPFLHRYQROLSSPYRDLRKGVVYPYTGQKWEGLGTDLVSIHPGPNVTVRANIAATE 120  
Db 106 hpflhryqrolsspyrdlrkgvvyvpytgqkwegeigtldvlsiphgpnvtvranaaite 165  
QY 121 SDKFFINGSNMEGILGLAYAEIARPDSDLEPFFDSLVKQTHVPNLFSLQLCGAGFPLNQS 180  
Db 166 sdkffingsnwegilglayaeiarpddslepfdslvkqthvpnlfsllqlcgaafplnqs 225  
QY 181 EVLASVCGSMIIGIDHSLYTGSWYPIRREYVEVITIVRVEINGODLKMCKEYNYDK 240  
Db 226 evlasvggsmiigidhslytgslywtpirrewyvevliivrveingdalkmckeynydk 285  
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIIKAASSTKFPDGFGLGQVLCWQAGTTPWNIFPVIS 300  
Db 286 sivdsgrtlnrlpkkvfeavksikaasstekfpgfwlgeqlvcwgaattpwlnifpvis 345  
QY 301 LYLGEVTNQSFRITILPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 360  
Db 346 lylmgevtnqsfritilpqgylrpvedvatsqddcykfaissstgtvmgavimegfyyv 405  
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDE 407  
Db 406 fdrarkrigrfavsachvhdefrtaavegpfvtilmedecgynipqtde 452  
RESULT 10  
AAE10643  
ID AAE10643 standard; Protein; 459 AA.  
XX  
AC AAE10643;  
XX  
DT 10-DEC-2001 (first entry)  
XX  
DE Human-Asp 2(a) protein with (His)6 tag and lacking TM domain.  
XX  
KW Human; aspartyl protease 2a; Asp2a; amyloid precursor protein; APP;  
KW Alzheimer's disease; AD; dementia; neurofibrillary tangle; gliosis;  
KW amyloid plaque; neuronal loss; proteolytic; neurotrophic; neuroprotective.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /label= Signal\_peptide  
FT Protein 22..459  
FT /note= "Mature Human-Asp2(a) deltaTM (His)6 protein"  
FT Misc-difference 214 /note= "Encoded by CAC"  
FT Misc-difference 454 /note= "Encoded by CAG"  
FT Misc-difference 455 /note= "Encoded by CAG"  
FT Misc-difference 456 /note= "Encoded by CAG"  
FT Misc-difference 457 /note= "Encoded by CAG"  
FT Misc-difference 458 /note= "Encoded by CAG"  
FT Misc-difference 459 /note= "Encoded by CAG"  
FT /note= "Encoded by CAG"  
XX  
PN GB2357767-A.  
XX  
XX 04-JUL-2001.  
XX  
PD 22-SEP-2000; 2000GB-0023315.  
XX  
XX 23-SEP-1999; 99US-0155493.  
PR

[illegible]



modulators of amyloid-beta peptide production, for use in designing therapeutics for the treatment and prevention of Alzheimer's disease, dementia, formation of amyloid plaques, neurofibrillary tangles, gliosis and neuronal loss. APP isoforms are also used in methods for identifying inhibitors and modulators of human Asp2 activity. The invention relates to a method for identifying agents that modulate the activity of human aspartyl protease Asp2. Amyloid-beta peptides obtained from APP are used as a means to screen in cellular assays for the inhibitors of beta- and gamma- secretase. Hu-Asp DNA fragments are useful as probes or primers in polymerase chain reactions (PCR). The probes are useful for detecting Hu-Asp nucleic acids in in vitro assays and in Northern and Southern blots. The present sequence is Human aspartyl protease 2a (Hu-Asp2a) deltaTM (His)6 protein which is obtained by deletion of C-terminal transmembrane domain and addition of a hexa-Histidine tag at the C-terminal end of Hu-Asp2a. Human Asp2a has beta-secretase activity.

Sequence 459 AA;

Query Match 100.0%; Score 2156; DB 22; Length 459;  
Best Local Similarity 100.0%; Pred. No. 2.5e-210;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSGFVEMVNLKSGQGYVEMTVGSPPTLNLVDTGSSNFAVGAAP 60  
Db 46 etdeepeepgrgsgfvmvnlrgskggyvemtvgspptlnllydtgssnfavgaa 105  
QY 61 HPFLHRYQRLSSTYRDLRKGVVPTQGWKEGELGTLVSIPIHGNVTVRANIAITE 120  
Db 106 hpflhryqrlsstdrldrgvvyptqgkwegelgtlvsiiphgnvtvranaiaite 165  
QY 121 SDRFFINGSNWEGILGAYAEIARPDLSLEFFDLSLVKQTHVPLNLSLQLCGAGFPLNQS 180  
Db 166 sdrffingsnwegilgayaearpdlsleffdsllvkqthvplnlsllqcgagfplnqs 225  
QY 181 EVLASVGSMTIGIDHSHTGSLWYTPIRREYVYVIVRVEINGDQKMDCKEYNYDK 240  
Db 226 evlasvgssmtigidhshtgslwytpirreyvyvivrveingdqkmdckeynydk 285  
QY 241 SIYDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLGELVCGWAGTTPWNIFPVIS 300  
Db 286 siydsgettlnlrpkkvfeaaavksiaasstekfpdglglvewaggttwnifpv 345  
QY 301 LYLMEVTNSFRITILPQYLRPVEDVATSQDDCYKFAISQSGTGVMGAVIMEGYVYV 360  
Db 346 lylmevtndsfritilpqylrpvedvatsqddcykfaissgstgtvmgavimegy 405  
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVGPFTLDMEDCGYNIPQTD 407  
Db 406 fdrarkrigfavsachvhdefrtaavgpftldmedcgynipqtde 452

RESULT 12  
AAU06617

ID AAU06617 standard; Protein; 459 AA.

AC AAU06617;

XX 24-OCT-2001 (first entry)

DT Human-pro-Asp 2(a) delta TM (His)6.

DE Human; Aspartyl protease; beta-secretase; neurotropic; ASP2;  
KW neuroprotective; amyloid protein precursor; APP; Alzheimer's disease;  
KW amyloid-beta; Abeta; Human-pro-Asp 2(a) delta TM (His)6; mutant; mutein.

XX Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers

FX Peptide 1..21

FT /label= Signal\_peptide

FT Protein 22..459

FT Misc-difference 214 /label= Mature\_Human\_pro\_Asp\_2(a)\_delta\_TM\_(His)6  
FT /note= "Encoded by CAC"  
FT Misc-difference 454..459  
FT /note= "Encoded by CAGCAGCAGCAGCAGCAG"  
FT Region 454..459  
FT /label= His\_tag  
FT /note= "Nickel binding region to aid purification"

PN WO200149098-A2.

XX 12-JUL-2001.

XX 09-MAY-2001; 2001WO-IB00798.

XX 09-MAY-2001; 2001WO-IB00798.

XX (BIEN/) BIENKOWSKI M J.

XX (GURN/) GURNEY M E.

XX (HEIN/) HEINRIKSON R L.

XX (PARO/) PARODI L A.

XX (YANR/) YAN R.

XX Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;

XX WPI; 2001-502549/55.

XX N-PSDB; AAS11531.

XX Novel purified polypeptide comprising fragment of mammalian aspartyl  
FT protease 2, lacking Asp2 transmembrane domain and retaining beta  
FT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
FT activity

XX Claim 149; Fig 12; 185pp; English.

XX The invention relates to a purified polypeptide comprising a fragment of  
CC mammalian aspartyl protease (Asp)2 protein which lacks the Asp2  
CC transmembrane domain and the Asp2 protein, and where the polypeptide and  
CC the fragment retain the beta-secretase activity of the mammalian Asp2  
CC protein. The invention also details polynucleotides for the Asp  
CC proteins and vectors expressing them, and a polypeptide (isoform of  
CC amyloid protein precursor (APP)) comprising the amino acid sequence of an  
CC APP or its fragment containing an APP cleavage site recognizable by a  
CC mammalian beta-secretase, and further comprising two lysine residues at  
CC the carboxyl terminus of the amino acid sequence of the mammalian APP or  
CC APP fragment. Also included in the invention are methods of identifying  
CC modulators or inhibitors of Asp2. Modulators and inhibitors of Asp2 are  
CC useful for treating Alzheimer's disease. APP is useful in methods for  
CC identifying inhibitors or modulators of human Asp2 activity and  
CC amyloid-beta (Abeta) peptide production. APP is also useful in designing  
CC therapeutics for the treatment or prevention of Alzheimer's disease.  
CC APP comprising the APP-Sw-beta-secretase peptide sequence (NLDA), which  
CC is associated with increased levels of Abeta processing is useful in  
CC assays relating the Alzheimer's research. The expression vector is useful  
CC for recombinantly expressing APP. Nucleic acids that hybridize to  
CC Asp oligonucleotides are useful as probes or primers. The probes are  
CC useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
CC Northern and Southern blots. The present sequence is Human-pro-  
CC Asp 2(a) delta TM (His)6 protein, which lacks the C-terminal  
CC transmembrane domain and has a His tag to aid purification.

XX Sequence 459 AA;

Query Match 100.0%; Score 2156; DB 22; Length 459;

Best Local Similarity 100.0%; Pred. No. 2.5e-210;

Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSGFVEMVNLKSGQGYVEMTVGSPPTLNLVDTGSSNFAVGAAP 60

Db 46 etdeepeepgrgsgfvmvnlrgskggyvemtvgspptlnllydtgssnfavgaa 105

QY 61 HPFLHRYQRLSSTYRDLRKGVVPTQGWKEGELGTLVSIPIHGNVTVRANIAITE 120

Db	106	hpfihryyqrlsstyrldlrkgyvpytgkgwgelgtdlvsiphgpnvtvranaiaite	165
QY	121	SDKFFINGSNWEGILGLAYAEIARPDLSLEPFFDSLQKTHVPLNLSLQICGAGFPLNQS	180
Db	166	sdkfingsnwegilglayaeiarpdlslepffdsivkqthvplnlsqicgagfplngs	225
QY	181	EVLASVGSMTIIGIDHSLYTGSLWTPPIRREWYEVIIIVRVEINGQDLKMDCKEYNYDK	240
Db	226	evlasvggsmiigidhslytgslytpirrewyevilvrveingqdlkmdckeynydk	285
QY	241	SIYDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFWLGEQLVCWQAGTTPWNIPFVIS	300
Db	286	sivdsgttnlrpkkvfeavksikaasstekfpgdglvcwgagttwnipfvis	345
QY	301	LYLMGEVTNQSFRITILPQOYLPRVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYVV	360
Db	346	lylmgevtnqsfrilitpqgylrpvedvatsqddcykfaissgstgtvmgavimegfyyv	405
QY	361	FDRAKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTD 407	
Db	406	fdarkrigfavsachvhdefrtaavegpfvtldmedcgynipqtde 452	
RESULT 13			
AAU07216			
ID	AAU07216 standard; Protein; 459 AA.		
XX	AAU07216;		
AC			
XX	24-OCT-2001 (first entry)		
DT			
XX	Human aspartyl protease 2a deltaTM (His)6.		
DE			
XX	Human; aspartyl protease 1; Asp-1; nootropic; neuroprotective;		
KW	aspartyl protease 2; Asp2; amyloid protein precursor; APP;		
KW	beta-secretase; Alzheimer's disease; HuAsp-2adeltaTM (His)6.		
XX	Homo sapiens.		
OS			
XX			
XX			
PH	Key Location/Qualifiers		
FT	Peptide	1..21 "Signal peptide"	
FT	/note= "Signal peptide"		
FT	Protein	22..459	
FT	/note= "Mature human aspartyl protease 2a deltaTM (His)6"		
FT	Misc-difference	454	
FT	/note= "Encoded by cag"		
FT	Misc-difference	455	
FT	/note= "Encoded by cag"		
FT	Misc-difference	456	
FT	/note= "Encoded by cag"		
FT	Misc-difference	457	
FT	/note= "Encoded by cag"		
FT	Misc-difference	458	
FT	/note= "Encoded by cag"		
FT	Misc-difference	459	
FT	/note= "Encoded by cag"		
XX	WO200149097-A2.		
PN			
XX	12-JUL-2001.		
PD			
XX	09-MAY-2001; 2001WO-IB00797.		
PF			
XX	09-MAY-2001; 2001WO-IB00797.		
PR			
XX	(BIEN/) BIENKOWSKI M J.		
PA	(GURN/) GURNEY M E.		
PA	(HEIN/) HEINRIKSON R L.		
PA	(PARO/) PARODI L A.		
XX	(YANR/) YAN R.		
PI	Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;		

XX	WPI; 2001-502548/55.			
DR	N-PSDB; AAS11716.			
XX				
PT	Novel purified polypeptide comprising fragment of mammalian aspartyl			
PT	protease 2, lacking Asp2 transmembrane domain and retaining beta			
PT	secretase activity of Asp2 useful for identifying inhibitors of Asp2			
PT	activity			
XX				
PS	Claim 149; Fig 12; 185pp; English.			
XX				
CC	The invention relates to a novel purified polypeptide comprising a			
CC	fragment of mammalian aspartyl protease 2 (Asp2) protein which lacks the			
CC	Asp2 transmembrane domain and the Asp2 protein, and where the polypeptide			
CC	and the fragment retain the beta-secretase activity of the mammalian Asp2			
CC	protein. Also included is an isoform of amyloid protein precursor (APP)			
CC	comprising the amino acid sequence of a APP or its fragment containing			
CC	an APP cleavage site recognisable by a mammalian beta-secretase, and			
CC	further comprising two lysine residues at the carboxyl terminus of the			
CC	amino acid sequence of the mammalian APP or APP fragment. The			
CC	polypeptides are used for assaying for modulators of beta-secretase			
CC	activity; identifying agents that inhibit the APP processing activity			
CC	of human Asp2 aspartyl protease (Hu-Asp2); identifying agents that			
CC	modulate the activity of Asp2; and for reducing cellular production of			
CC	amyloid beta (Abeta) from APP. Agents identified by the above methods			
CC	are useful for treating Alzheimer's disease; and for identifying			
CC	modulators of amyloid-beta (Abeta) peptide production, for use in			
CC	designing therapeutics for the treatment or prevention of Alzheimer's			
CC	disease. Probes and primers derived from Asp nucleic acid sequences			
CC	are useful for detecting Hu-Asp nucleic acids in in vitro assays and in			
CC	Northern and Southern blots. The present sequence represents the amino			
CC	acid sequence of human Asp-2a delta TM (His)6 construct which has			
CC	a 6 histidine tag and lacks the transmembrane domain. This construct was			
CC	used for expression and purification of human Asp2a in insect cells.			
XX				
SQ	Sequence 459 AA;			
	Query Match	100.0%;	Score 2156;	DB 22; Length 459;
	Best Local Similarity	100.0%;	Pred. No. 2.5e-210;	
	Matches 407;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY	1	ETDEEPEEPRGRGSFVEMVDNLKCGSGQGYVVTQGWEGELGTLVSIPIHGPNTVVRANIAAITE	60	
Db	46	etdeeepeepgrgsfvmvndnlrgksqgyvvtvsgpqtlnilvdtgssnfavgaap	105	
QY	61	HPFLHRYQRLSSTYRDLRKGVVYPTQGWEGELGTLVSIPIHGPNTVVRANIAAITE	120	
Db	106	hpfihryyqrlsstyrdlrkgyvpytgkgwgelgtdlvsiphgpnvtvranaiaite	165	
QY	121	SDKFFINGSNWEGILGLAYAEIARPDLSLEPFFDSLQKQTHVPLNLSQLCGAGFPLNQS	180	
Db	166	sdkffingsnwegilglayaeiarpdlslepffdsllvkqthvplnlsqlcgagfplnqs	225	
QY	181	EVLASVGSMTIIGIDHSLYTGSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNYDK	240	
Db	226	evlasvggsmiigidhsllytgslywtpirrewyevliivrvellingqdlkmdckeynydk	285	
QY	241	SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFWLGEQLVCWQAGTTPWNIPFVIS	300	
Db	286	sivdsgttnlrpkkvfeaaavksikaasstekfpgdglvcqlvcwgagttwnipfvis	345	
QY	301	LYLMGEVTNQSFRITILPQOYLPRVEDVATSDQCYKFAISQSSTGTVMGAVIMEGFYVV	360	
Db	346	lylmgevtnqsfrilitlpqgylrpvedvatssdqcykfaissgstgtvmgavimegfyyv	405	
QY	361	FDRAKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTD 407		
Db	406	fdrarkrigfavsachvhdefrtaavegpfvtldmedcgynipqtd 452		
RESULT 14				
AAE02595				



Wed Aug 7 10:38:04 2002

us-09-724-571-58.rag

CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
CC present sequence represents a human beta-secretase enzyme fragment.

XX	Sequence	460 AA;
SQ		
Query Match		100.0%; Score 2156; DB 21; Length 460;
Best Local Similarity		100.0%; Pred. No. 2.5e-210;
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	ETDEPEPEGRGSEFVEMVNLNKGSGGYVEMTVGSPPTLNILVDTGSSNFVAGAAP 60
Db	46	etdeepeepgrgsfvevnmvnlrgksqgyvemtvgspptlnilvdtgssnfargaa 105
QY	61	HPFLHRYYQRLSSYTRDLRKGVVYPYTOGKWEGELGTDLVSIPIHGPNTVVRANIAITE 120
Db	106	hpflhryyqrlsstyrdlrgkvvypytgkwegeltdivsiiphgpnvtvranaiaite 165
QY	121	SDKFFINGSNWEGILGLAYAEIARPDLSLEPFEDSLVKQTHVPNLFSLQLCGAGFPLNQS 180
Db	166	sdkfingsnwegilglayaeiarpddslepfidslvkqthvpnlfsiqcagfplnqs 225
QY	181	EVLASVGSMIIGGDHSLYTGSLWYTPIRREWYIEYIIVRVEINGQDLKMDCKEYNKYDK 240
Db	226	evlasvggsmliggidhslytgslywtpirrewyieyliivrveingqdlkmdckeynydk 285
QY	241	STVDSGTTNLRPLPKVFEAAVKSIAASSTKFFDGFWLGEOLVCWQAGTTPWNIFPVIS 300
Db	286	sivdsgettlnrlpkkvfeavksikaasstekffdgfwlgeqlvcwdagttppwnifpv 345
QY	301	LYLMGEVTNQSFRTILPQQYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYVV 360
Db	346	lylmgevtncsfritilpqylrpvedvatsqddcykfalsqsstgtvmgavimegfyyv 405
QY	361	FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDTE 407
Db	406	fdrarkrigfavsachvhdefrtaavegpfvtldmedcgynipqtd 452

Search completed: August 7, 2002, 09:16:15  
Job time: 160 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:14:17 ; Search time 32.31 Seconds  
(without alignments)  
1210.410 Million cell updates/sec

Title: US-09-724-571-58  
Perfect score: 2156  
Sequence: 1 ETDEPEPGRGRGFVEMVD.....GPFVLDMDGYNIPQTD 407

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: PIR1:\*
- 2: PIR2:\*
- 3: PIR3:\*
- 4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2156	100.0	501	2 A50900	aspartic proteinase
2	308	14.3	384	2 JC7574	pepsinogen A - Afr
3	308	14.3	385	2 JC7575	pepsinogen A - bul
4	306	14.2	387	2 B38302	pepsin (EC 3.4.23)
5	303	14.1	383	2 JC7573	pepsinogen C - Afr
6	302.5	14.0	388	1 S19682	pepsin A (EC 3.4.23)
7	301	14.0	382	1 PECH	pepsin A (EC 3.4.23)
8	299.5	13.9	396	2 A34401	pepsin E (EC 3.4.23)
9	298.5	13.8	383	2 A41443	pepsin (EC 3.4.23)
10	298.5	13.8	384	2 A39314	gastricsin (EC 3.4.23)
11	296	13.7	387	2 B38302	pepsin (EC 3.4.23)
12	296	13.7	391	2 A43356	pepsinogen E (EC 3.4.23)
13	295.5	13.7	412	1 KHUD	pepsinogen C - Afr
14	295	13.7	387	2 D38302	pepsin (EC 3.4.23)
15	289.5	13.4	388	1 S19684	pepsin A (EC 3.4.23)
16	287.5	13.3	444	2 T24204	hypothetical prote
17	287	13.3	407	1 KHRD	pepsin D (EC 3.4.23)
18	286	13.3	387	2 B38302	pepsin (EC 3.4.23)
19	285.5	13.2	398	2 S66465	pepsin A (EC 3.4.23)
20	285	13.2	398	2 I51185	pepsin D (EC 3.4.23)
21	283.5	13.1	388	1 PEMQAR	pepsin A (EC 3.4.23)
22	282.5	13.1	388	1 PSHU	pepsin A (EC 3.4.23)
23	281.5	13.1	381	1 CMSHB	pepsin A (EC 3.4.23)
24	281.5	13.1	388	2 A30142	chymosin (EC 3.4.23)
25	281.5	13.1	388	2 PEMQAJ	pepsin A (EC 3.4.23)
26	279.5	13.0	388	2 B30142	pepsin A (EC 3.4.23)
27	279.5	13.0	410	1 KMSD	pepsin A (EC 3.4.23)
28	278.5	12.9	385	1 PEFG	pepsin A (EC 3.4.23)
29	278	12.9	387	2 JC7245	pepsinogen A - com

30	277.5	12.9	380	2 I47176	chymosin (EC 3.4.23)
31	277.5	12.9	396	2 S36865	cathepsin E (EC 3.4.23)
32	276	12.8	389	2 JE0371	pepsin C (EC 3.4.23)
33	273.5	12.7	381	1 CMBO	chymosin (EC 3.4.23)
34	270.5	12.5	377	1 PEMQJ	pepsin (EC 3.4.23)
35	270.5	12.5	389	2 A38302	aspartic proteinase
36	270	12.5	376	2 I45856	aspartic proteinase
37	268.5	12.5	344	1 KHFGD	cathepsin D (EC 3.4.23)
38	267.5	12.4	381	2 JC7247	prochymosin - com
39	266	12.3	380	2 S03433	candidapepsin (EC 3.4.23)
40	266	12.3	405	2 A25379	saccharopepsin (EC 3.4.23)
41	264	12.2	396	2 T47207	aspartic proteinase
42	263.5	12.2	388	2 JC7246	pepsinogen C - com
43	262.5	12.2	394	2 B43356	gastricsin (EC 3.4.23)
44	261.5	12.1	387	2 A45117	aspartic proteinase
45	261.5	12.1	388	2 A29937	gastricsin (EC 3.4.23)

ALIGNMENTS

RESULT 1

A50900

N:Alternate proteinase (EC 3.4.23.-) BACE precursor - human  
N:Alternate names: beta-secretase; beta-site APP cleaving enzyme  
C:Species: homo sapiens (man)  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 11-May-2000  
C:Accession: A50900

R:Vassar, R.; Bennett, B.D.; Babu-Khan, S.; Kahn, S.; Mendiaz, E.A.; Denis, P.; Teplö  
M.A.; Biere, A.L.; Curran, E.; Burgess, T.; Louis, J.C.; Collins, F.; Treanor, J.; Ro  
Science 286, 735-741, 1999  
A:Title: beta-Secretase cleavage of Alzheimer's amyloid precursor protein by the tran  
A:Reference number: A50900; MUID:20002972  
A:Note: submitted to GenBank, September 1999  
A:Accession: A50900  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-501 <VAS>  
A:Cross-references: GB:AF190725; NID:96118538; PIDN:AAF04142.1; PID:96118539  
C:Genetics:  
C:Gene: BACE  
C:Superfamily: beta-secretase  
C:Keywords: Alzheimer's disease; aspartic proteinase; brain; glycoprotein; hydrolase;  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-45/Domain: signal sequence #status predicted <PRO>  
F:46-501/Product: acid proteinase BACE #status predicted <MAT>  
F:461-477/Domain: transmembrane #status predicted <TRN>  
F:93,289/Active site: Asp #status predicted  
F:153,172,223,354/Binding site: carboxylate (Asn) (covalent) #status predicted  
F:330-380/Disulfide bonds: #status predicted

Query Match 100.0%; Score 2156; DB 2; Length 501;  
Best Local Similarity 100.0%; Pred. No. 1.6e-173;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ETDEPEPGRGRGFVEMVDNLKRGSGQGYVVTGSPPTNLVDTGSSNFAVGAAP	60
Db	46	ETDEPEPGRGRGFVEMVDNLKRGSGQGYVVTGSPPTNLVDTGSSNFAVGAAP	105
Qy	61	HPFLHRYQRLSTYRDLRKGYYVVTGQKWEGLGTDLVSIPIHGNVTVRANIAITE	120
Db	106	HPFLHRYQRLSTYRDLRKGYYVVTGQKWEGLGTDLVSIPIHGNVTVRANIAITE	165
Qy	121	SDKFFINGSNWEGTLGLAYAEIARPDLSLPPFDLSLVKQTHVPLNLSLQCGAGFP	180
Db	166	SDKFFINGSNWEGTLGLAYAEIARPDLSLPPFDLSLVKQTHVPLNLSLQCGAGFP	225
Qy	181	EVLASVGSMTIGIDHSLTGSLSWYTPIRREWYEVIIIVRVEINGDLKMDCKEYNDK	240
Db	226	EVLASVGSMTIGIDHSLTGSLSWYTPIRREWYEVIIIVRVEINGDLKMDCKEYNDK	285
Qy	241	SIYDSGTNLRPKKVEAAVKSIKAASSTKFPDGFGLVCMQAGTTPWNIFPVIS	300

Wed Aug 7 10:38:05 2002

```
|||||
286 SIVDSGTTNLRPKVFEAAVSKAASSTKFPDGFGLVQWAGTTWNIPFVLIS 345
QY 301 LYLMGEVWNSFRITILPQOYLRPVEDVATSDDCYKFAISQSSSTGTVMGAVIMEGFYV 360
|||||
346 LYLMGEVWNSFRITILPQOYLRPVEDVATSDDCYKFAISQSSSTGTVMGAVIMEGFYV 405
QY 361 FDRARRKIGFAVSACHVHDFRTAAVEGPFVTLDMEDCGYNIPQTD 407
|||||
406 FDRARRKIGFAVSACHVHDFRTAAVEGPFVTLDMEDCGYNIPQTD 452
Db

RESULT 2
JC7574
pepsinogen A - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
C:Accession: JC7574; PC7119
R;Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens
A:Reference number: JC7573; MUID:21064922; PMID:11134969
A:Contents: Stomach
A:Accession: JC7574
A:Molecule type: mRNA
A:Residues: 1-384 <IKU>
A:Cross-references: DDBJ:AB045380
A:Accession: PC7119
A:Molecule type: protein
A:Residues: 16-35;57-76 <IK2>
A:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like
C:Genetics:
A:Gene: Pga
C:Superfamily: pepsin
C:Keywords: stomach; zymogen

Query Match 14.3%; Score 308; DB 2; Length 384;
Best Local Similarity 27.5%; Pred. No. 4.8e-18;
Matches 99; Conservative 60; Mismatches 133; Indels 68; Gaps 16;
QY 30 YVEMTVGSPQPTNLILVDTGSSNFAVGAAPHPFL-----HRYRQLSSTYRDLRKG 82
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 YGTISIGTPPQSFVTFIDTGSANLWV---PSVYCSQACSNHNFNPQSSFTQATNP 128
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 83 VYPTQGWKEGELGDLVSIHPGPNVTVRANIAATESDK-FFINGSNWEGILGLAYAE 141
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 129 VSIQYGTGSMGFLGYDTLQV---GNIQISNQMFGLSESEPGSFLYSPFDGILGLAFPS 185
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 142 IARPDLSLEPFDSLVKQTHVP-NLFSQLQCGAGFPINQSEVLASVGGSMITGGIDHSLY 200
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 186 IA--SSQATPVFDNMWSQGLIPQNLFSVYSSDG-----QTGSYVLFGGVDNSY 233
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 201 TGSLSWYTPIRREWYEVIIIVRVEINGDQL--KMDCKEYNDKSIDSGTTNLRPKKVE 258
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 234 SGLSNWVPLAETWYQITLDSVINGVIACSSC-----QAIVDTGSLMTGPSTPI- 286
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 259 AAVKSIKAASSTKFPDGFGLVQWAGTTWNIPFVLSLYLMGEVWNSFRITIL 317
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 287 ANIQNYTGASQDSN-----GQVINCNNISNMPTIVE-----TIN 321
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 318 PQQY-LRPVEDVATSDDCYK-FAISQSSSTGT-----VMGAVIMEGFYVDFDRARRKIGFA 371
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 322 GVQYPLPSAVVRNQCGSSGSGFQAMNLPNTNSGDLWILGDVFIQYFTVDFDRANNYVAIA 381
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db

RESULT 3
JC7575
pepsinogen A - bullfrog
C:Species: Rana catesbeiana (bullfrog)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
C:Accession: JC7575
R;Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens
A:Reference number: JC7573; MUID:21064922; PMID:11134969
A:Contents: Stomach
A:Accession: JC7575
A:Molecule type: mRNA
A:Residues: 1-385 <IKU>
A:Cross-references: DDBJ:AB045376
A:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like
C:Genetics:
A:Gene: Pga
C:Superfamily: pepsin
C:Keywords: stomach; zymogen

Query Match 14.2%; Score 306; DB 2; Length 387;
Best Local Similarity 27.1%; Pred. No. 7.2e-18;
Matches 98; Conservative 67; Mismatches 131; Indels 66; Gaps 15;
QY 30 YVEMTVGSPQPTNLILVDTGSSNFAVGA-----AAPHPFLHRYRQLSSTYRDLRKG 85
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 75 YGTISIGTPPQSFVTFIDTGSANLWVPSVYCSQACSNHNFNPQSSFTQATNPVSI 134
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 86 PYTGKWEGELGDLVSIHPGPNVTVRANIAATESDK-FFINGSNWEGILGLAYAEI 142
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db

J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinog
A:Reference number: JC7573; MUID:21064922; PMID:11134969
A:Contents: Stomach
A:Accession: JC7575
A:Molecule type: mRNA
A:Residues: 1-385 <IKU>
A:Cross-references: DDBJ:AB045376
A:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like
C:Genetics:
A:Gene: Pga
C:Superfamily: pepsin
C:Keywords: stomach; zymogen

Query Match 14.3%; Score 308; DB 2; Length 385;
Best Local Similarity 27.8%; Pred. No. 4.8e-18;
Matches 99; Conservative 64; Mismatches 133; Indels 60; Gaps 15;
QY 30 YVEMTVGSPQPTNLILVDTGSSNFAVG---AAPHPLHRYRQLSSTYRDLRKG 85
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 73 YGTISIGTPPQSFVTFIDTGSANLWVPSVYCSQACSNHNFNPQSSFTQATNPVSI 132
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 86 PYTGKWEGELGDLVSIHPGPNVTVRANIAATESDK-FFINGSNWEGILGLAYAEI 144
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 133 QYGTGSMGFLGYDTQV---GNIQITNQIFGLSQSEPGSFLYSPFDGILGLAFPSLA- 188
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 145 PDSLEPFDSLVKQTHVP-NLFSQLQCGAGFPINQSEVLASVGGSMITGGIDHSLYTGS 203
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 189 -SSQATPVFDNMWNOGLIPQDLFSVYSSOG---QS-----GSFVLFGGVDTSYTGN 237
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 204 LWTPIRREWYEVIIIVRVEINGDQLK--DCKEYNDKSIDSGTTNLRPKKVEFAV 261
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 238 LNWVPLAETWYQITLDSVINGVIACSSC-----SAIVDTGTSLLAGP-----STPI 287
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 262 KSIKAASSTKFPDGFGLVQWAGTTWNIPFVLSLYLMGEVWNSFRITILPQOY 321
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 288 ANIQYIGANODSNGQV---INCNNISNMPTVVF-----TINGVOY 326
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 322 LRPVED-VATSDDC---YKFAISQSSSTGT---VMGAVIMEGFYVDFDRARRKIGFA 371
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 327 PLPASVVRQSQSCTSGFQAMNLPNTNSGDLWILGDVFIQYFTVDFDRANNYVAMA 382
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db

RESULT 4
B38302
pepsin (EC 3.4.23.-) II-1 precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 23-Feb-1997
C:Accession: B38302
R;Kageyama, T.; Tanabe, K.; Koiwai, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu
A:Reference number: A38302; MUID:91009127
A:Accession: B38302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:M59235; GB:J05638
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 14.2%; Score 306; DB 2; Length 387;
Best Local Similarity 27.1%; Pred. No. 7.2e-18;
Matches 98; Conservative 67; Mismatches 131; Indels 66; Gaps 15;
QY 30 YVEMTVGSPQPTNLILVDTGSSNFAVG---AAPHPLHRYRQLSSTYRDLRKG 85
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 75 YGTISIGTPPQSFVTFIDTGSANLWVPSVYCSQACSNHNFNPQSSFTQATNPVSI 134
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 86 PYTGKWEGELGDLVSIHPGPNVTVRANIAATESDK-FFINGSNWEGILGLAYAEI 142
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db
```

[illegible]

RESULT	8	
A3401	Cathesin E (EC 3.4.23.34) precursor - human	22-Jun-1999
A3401	Cathepsin E; Homo sapiens (man)	22-Jun-1999
C.Crpe:	22-Jun-1990 sequence revision #32467; A34643	22-Jun-1999
C.Crpe:	22-Jun-1990 sequence revision #32467; A34643	22-Jun-1999
C.Accession:	A42038 A3401: S35663; S34467; B34643	
R.Azuma,	T.; Liu, W.; Vander Laan, D.J.; Bowcock, A.M.; Taggart, R.T.	
I Biol Chem-	267. 1609-1614, 1992	

	Query Match	13.9%;	Score	299.5;	DB 2:	Length	396;			
	Best Local Similarity	25.9%;	Pred. No.	2.6e-17;						
	Matches	100;	Conservative	65;	Mismatches	148;	Indels	73;	Gaps	16;
QY	3	DEEPEEPGRGGSFVEMVDNLKRGSGGYVEMTVGPSPQTLLNLDVTGSSNPVACA----	58							
		:     :	:	:	:	:	:	:	:	
Dd	63	DQSAAKEP-----LNYLD-----MEYFGTISGGPPQNFTVFDTGSSNLWPSVYCT	110							
		:     :	:	:	:	:	:	:	:	
QY	59	APHPLPHRYQRQLSSTYRDLRKGVYPPTYQGKNWEGELGTDLVSIPHGPNVTVRAIAAI	118							
		:     :	:	:	:	:	:	:	:	
Dd	111	SPACKTHSRFPQSOSTYSVPQGSFSIQYGTGSLGIAGADQSV--EGITVVYQQGFGEV	169							
		:     :	:	:	:	:	:	:	:	
QY	119	TESDRFFINGSNWEGILGLVAETARPDDLSLEPFEDSKVKQTHVPNFLSLQLCGAGFPFL	178							
		:     :	:	:	:	:	:	:	:	
Dd	170	TPGQDTFVD-AFFGDGILGLGPSIA---VGGVTPVFQDNMAQAQ-----NLVLDPMFVSVMSSN	222							
		:     :	:	:	:	:	:	:	:	
QY	179	QSEVLASVGGSIMIGGDHDSLYTGSLMWTPTRRWYEYEVILLVRVEINGDQLMKDCKEYNY	238							



Db 223 PE---GGAGSELIFGGYDHSFSGSLNWPVTKQAYQIALDNIQVGG---TMFCSB--G 275  
Qy 239 DKSIVDSGTTNLRPKKVFEEAAVKSIAASTKFPDGFWMGEOLVQWQAGTTPWNIFPV 298  
Db 276 COAIVDTGTSLSITGSDKIKOLQNAAGP-----VDGEYAVE-----CANLVMPD 322  
Qy 299 ISLYLMGEVNTQSFRIITLPOQLRPVEDVATSDCCYKFAISQSSTG----- 346  
Db 323 VTFTING-----VPYTLSPAY---TLDFVDMQFC-----SSGFQGLDIHPAGP 366  
Qy 347 -TVMGAVIMEGFYVDFRANKRIGFA 371  
Db 367 LWILGDVFIQFYSVDFDRGNRVGLA 392

RESULT 9  
A41443  
pepsin (EC 3.4.23.-) precursor, embryonic - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 05-Jun-1992 #sequence\_revision 05-Jun-1992 #text\_change 21-Jul-2000  
C:Accession: A41443  
R:Hayashi, K.; Agata, K.; Mochii, M.; Yasugi, S.; Eguchi, G.; Mizuno, T.  
A:Title: Molecular cloning and the nucleotide sequence of cDNA for embryonic chicken pepsin  
A:Reference number: A41443; MUID:88227903  
A:Accession: A41443  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-383 <HAY>  
A:Cross-references: GB:D00215; NID:g2760810; PIDN:BAA00153.1; PID:g222853  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; hydrolase; protein digestion

Query Match 13.8%; Score 298.5; DB 2; Length 383;  
Best Local Similarity 25.2%; Pred. No. 3e-17;  
Matches 90; Conservative 75; Mismatches 125; Indels 67; Gaps 14;  
Qy 30 YVEMTVGSPPTNLILVDTGSSNFAVGA-----APHPFLHRYQRLSSYRDLRKGVYV 85  
Db 76 YGTISGTPPDFTVDFDTGSSNLWPSVCTSPACQSHQMFNPSSSTYKSTGQNLIS 135  
Qy 86 PYTGKWEGLGTLVSLPHGPNVTVRANIAAITESSDKFFETNGSNWEGILGLAYAEATRP 145  
Db 136 HYGTGDMGTGCDVTVASLMDTNQLFGLST-SEPGQFVY-VKFDGILGLGYPSLAA- 192  
Qy 146 DDLPEPFDLSLVKQTHV-PNLFSLQLCAGFPLNQSEVLASVSGSMIIGGIDHSLYTGSL 204  
Db 193 -DGITPVFDNNVNSLLEQNLFVYLS-----REPMSMVVFGGIDBSYFTGSI 240  
Qy 205 WYTPIRREWYEVILVRVEINGQDL--KMCKEYNDKSIYDSGTTNLRPKKVFEEAAVK 262  
Db 241 NWIPVSGYVQWISMDSIYNKQBIACSSG-----QALIDGTSLVAGPASDINDIQS 294  
Qy 263 SIKAASTEKPPDGFWMGEOLVQWQAGTTPWNIFPVISL-----YLMGEVNTQSFRIITLP 318  
Db 295 AVGANQNT-----YGEYSV-----NCSHILAMDVVVFTGGI----- 326  
Qy 319 QOYLRPVEDVA-----TSQDCCYKFAISQSSTGTVMGAVIMEGFYVDFRANKRIGFA 371  
Db 327 -QY--PVPALAYTEQCGQCTMCSFQNSSADLWILGDVFIQFYSVDFDRANNRVGLA 380

RESULT 10  
A39314  
gastricsin (EC 3.4.23.3) precursor - bullfrog  
C:Species: Rana catesbeiana (bullfrog)  
C:Date: 19-Jun-1992 #sequence\_revision 19-Jun-1992 #text\_change 22-Jun-1999  
C:Accession: A39314  
R:Yakabe, E.; Tanji, M.; Ichinose, M.; Goto, S.; Miki, K.; Kurokawa, K.; Ito, H.; Kageya  
J. Biol. Chem. 266, 22436-22443, 1991  
A:Title: Purification, characterization, and amino acid sequences of pepsinogens and pep

A:Reference number: A39314; MUID:92042186  
A:Accession: A39314  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-384 <YAK>  
A:Cross-references: GB:M73750; NID:g213687; PIDN:AAA49530.1; PID:g213688  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; hydrolase; protein digestion

Query Match 13.8%; Score 298.5; DB 2; Length 384;  
Best Local Similarity 25.5%; Pred. No. 3e-17;  
Matches 100; Conservative 61; Mismatches 142; Indels 89; Gaps 15;  
Qy 14 SFVEMVDNLCKSGGYYVEMTVGSPPTNLILVDTGSSNFAVGAAPHPFL-----HR 66  
Db 51 NFATAFEPLANMDSYGEISIGTPPQNLFLVDFDTGSSNLWV---PSTYCSQSQACTNHP 107  
Qy 67 YYQRLSSTYRDLRKGVYVPTQKWEGLGTLVSLPHGPNVTVRANIA-----AI 118  
Db 108 QFNPSQSSSYSSNQOQFSLQYGTGSLNGILGYDTVQI-----QNTAISQOEFGLSV 158  
Qy 119 TESDKFFETNGSNWEGILGLAYAEATRPDSDLEPFFDLSVKQTHVPN-LFSLQLCAGFPL 177  
Db 159 TEPQTNFY-AQFDGILGLAYPSIA--EGGATTVMQGMIOQNLINQPLFAFYLSGQONSQ 215  
Qy 178 NQSEVLASVSGSMIIGIDHSLYTGSLWYTPIRREWYEVILVRVEINGQD---LKMCK 234  
Db 216 N-----GGEAVFGVDQNYSGQIYWTVPVTSYTWQIGQGFVNGQATGWCSSQGC- 266  
Qy 235 EYNVDKSIYDSGTTNLRPKKVFEEAAVKSIAASTKFPDGFWMGEOLV-CWQAGTTPW 293  
Db 267 -----QGIVDTGTSLLTAPQSVESSLMQSIGAQDQN-----QYAVSCSNQSLSPT 313  
Qy 294 NIFPVI-----SLYLMGEVNTQ---SFRITLPQOYLRPVEDVATSDCCYKFAISQ 342  
Db 314 ISFTISGVSFPPLPSAYVLAQNSGYCTIGIMTYLPSQNGQL----- 356  
Qy 343 SSTGTWAGVIMEGFYVDFRANKRIGFAVSA 374  
Db 357 ----WILGDVFLRQYYSYVDLGNQNVGFAAAA 384

RESULT 11  
C38302  
pepsin (EC 3.4.23.-) II-2/3 precursor - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 14-Jun-1991 #sequence\_revision 14-Jun-1991 #text\_change 23-Feb-1997  
C:Accession: C38302  
R:Kageyama, T.; Tanabe, K.; Koiwai, O.  
J. Biol. Chem. 265, 17031-17038, 1990  
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu  
A:Reference number: A38302; MUID:91009127  
A:Accession: C38302  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-387 <KAG>  
A:Cross-references: GB:J05638  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 13.7%; Score 296; DB 2; Length 387;  
Best Local Similarity 26.9%; Pred. No. 5e-17;  
Matches 97; Conservative 63; Mismatches 135; Indels 66; Gaps 13;  
Qy 30 YVEMTVGSPPTNLILVDTGSSNFAVGAAPHPF-----LHRYQRLSSYRDLRK 82  
Db 75 YFGTISGTPPDFTVDFDTGSSNLWV---PSTYCSLALCHKRNPEDSSTYQGTSET 131  
Qy 83 VVVPYTGKWEGLGTLVSLPHGPNVTVRANIAAITESSDKFFETNGSNWEGILGLAYAEI 142  
Db 132 LSITYGTGSMTGILGYDTVKGVSIEDTNQIFGLSKTSPSLTFLF--APFDGILGLAYPSI 189



A;Description: limited specificity endopeptidase  
A;Pathway: intracellular protein degradation

C:Keywords: aspartic proteinase; glycoprotein; hydrolase; lysosome; protein degradation  
F:1-60/Domain: signal sequence #status predicted <SIG>  
F:21-64/Domain: propeptide #status predicted <PRO>  
F:65-162,169-410/Product: cathepsin D #status experimental <MAT>  
F:167,329-356/Region: phosphotransferase recognition  
F:167,320,110-117,286-290,329-366/Disulfide bonds: #status experimental  
F:37,295/Active site: Asp #status experimental  
F:134,263/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 13.7%; Score 295.5; DB 1; Length 412;  
Best Local Similarity 28.5%; Pred. No. 6e-17;  
Matches 105; Conservative 58; Mismatches 146; Indels 59; Gaps 15;

**Qy** 30 YYVEMTVGSPQTINILVDVTGSSNFAVGAAHPFL-----HRYYQRQLSSTYRDLRKGV 83  
|| | : ||| : ||||| : | | | |  
**Db** 79 YYGEIGITPPQCFTVDFDTGSSNLWVPSTHCKLLDIACWIHKYNSDKSSTYYKNGTSF 138

**QY** 84 YVPYTQGWEGELGTDLVSLP-----HGPNTVTRANIAAITESDKFFINGSNWEGI 134  
: | | | | | : | | | | :  
**Dd** 139 DIHYGSSLSGYLSQDTSVPCQSASSASALGGVKVEROVFGATKQPCGTETAAKEFCGI 199

[illegible]

194 QY GIDHSLYTGSLWTPPIRREYYEYVIRVEI - NGQDLKMDCKEYNDKSIYDVGTTNRL 252  
 249 Db GTDSKVVYKGGT SVI NUNMPG A VEGTGVY DQYHVV GQVTL  
 -----SKDFDAQFGGELMLG 248

**QY**      253 PKKVFEEAAVKSIAKAASSTEKEFPDGFWLGOLV-CMQAGTTPWNFFVISLYLMGEVTNQS 311  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
**DH**      304 DDDGTTGGCTCKLSTGLSTLT--CPE--GCCEAIVDTGTSLMVG 303

304 PVDEVRELQRAIGAVPLIQ-----GEMNIPCEKVST-----LPAITLKLGG----KG 346  
 312 FRITILPQOYLRPVEDVATSDDCYKFAISQ-----SSTGTVMGAVTMEGFYVVFEDARK 366

DD 34/ YKLS--PEDYTLKVSQAGKTL--CLSGFGMGMDIPPSGGLWILGDVFIGRYTTFVDRDNN 402

QY 367 RIGFAVSA 374  
|:|:|:|

```
DB 403 RVGFAEAA 410
RESULT 14
```

D38302  
pepsin (EC 3.4.23.-) II-4 precursor - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 14-Jun-1991 #sequence revision 20-Sep-1991

C:Accession: D38302  
R:Kageyama, T.; Tanabe, K.; Koizumi, O.  
J. Biol. Chem. 265, 17031-17038, 1990  
A:Title: Structure and development of the  
C:Accession: D38302  
R:Kageyama, T.; Tanabe, K.; Koizumi, O.  
J. Biol. Chem. 265, 17031-17038, 1990  
A:Title: Structure and development of the

A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-387 <KAG>  
A:Cross-references: GB:M59235; GB:J05638  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match	13.7%;	Score 295;	DB 2;	Length 387;
Best Local Similarity	26.1%;	Pred. No. 6.1e-17;		
Matches 97; Conservative	65;	Mismatches 123;	Indels 86;	Gaps 14;

30 YVVENTGSPQTINILVDTGSSNFVGAAPHF-----LHRYQRLSSTYRDLRKG 82  
 75 YGTISITPPQDQTVTFDTGSSNLW---PSTYCSLALHKKRFPDSDSTYOGTSET 131

QY	83	YYVPYTGKQWEGELGTLVLSIPHGPNVTVRANIAAITESDKFF-----INGSNWE	132
Db	132	LSITYGTSGMTGLIGYDV-----KVGSIETNQIFGLSKTBPGTLFLFAPED	179
QY	133	GILGLAYAEATARPDDSLPEPFDLSLVKQTHV-PNLFSQLCCGAGFPLNQSEVLASVCGSMI	191
Db	180	GILGLAYPSISSSDAT--PVFDNNWNEGLVQDLFSVYSSDD-----EGKSLVM	227
QY	192	IGGIDHSLYTGSLSWYPTIRREWYEVIIVRVEINGODLKM--DCKEYNYDKSIVDSGTNN	249
Db	228	FGGIDSYTYTGSLSWNPVSVSEGYWQITMDSVINGETIACADSC-----QAIVDTGTSL	281
QY	250	LRLPKVFEEAAVKSIIKAASSTEFKPDGFWLGEOLV-CWQAGTTPWNIFFVLSYLMGEVT	308
Db	282	LTGP-----TSAISNIQSYICASK-----NULLGENVISCASIDSLPDIVF-----	321
QY	309	NQSFRTILPQOYLRPVEDVATSDQCYKFAISQSTGT-----VMGAVIMEGFYVVPD	362
Db	322	-----TINGIQPLPASAIKEDDDCTSGLEGMNVDTYTGELMWLGDVFIRQYTFVD	375
QY	363	RARKRIGFAVS	373
Db	376	RANNOLGLAA	386

RESULT 15

S19684

Pepsin A (EC 3.4.23.1) 2/3 precursor - Japanese macaque  
N; Alternate names: pepsinogen A isozyme 2/3  
C; Species: Macaca fuscata (Japanese macaque)  
C; Date: 22-Nov-1993 #sequence\_revision 19-Oct-1995 #text\_change 18-Jun-1999  
C; Accession: S19684; S16064  
R; Kageyama, T.; Tanabe, K.; Koikai, O.  
Eur. J. Biochem. 202, 205-215, 1991  
A; Title: Development-dependent expression of isozymogens of monkey pepsinogens and st  
A; Reference number: S19681; MUID:92037645  
A; Accession: S19684  
A; Molecule type: mRNA  
A; Residues: 1-388 <KAG>  
A; Cross-references: EMBL:X59755; NID:G38068; PIDN:CAA42427.1; PID:G38069  
A; Note: parts of sequence, including amino ends of pepsinogen and activation intermed  
C; Comment: It could not be determined if this sequence represents isozyme 2 or 3, whi  
in by 4 months of age.  
C; Comment: Although two-step activation is observed, activation is predominantly a o  
C; Superfamily: pepsin  
C; Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein di  
F.1-15/Domain: signal sequence status predicted <SIG>  
F.16-388/Product: pepsinogen A 2/3 #status experimental <PPT>  
F.16-62/Domain: activation peptide #status experimental <APT>  
F.63-388/Product: pepsin A 2/3 #status experimental <ENZ>  
F.40-41/Cleavage site: Asp-Phe (pepsin) #status experimental  
F.62-63/Cleavage site: Leu-Ile (pepsin) #status experimental  
F.94, 277/Active site: Asp #status predicted  
F.107-112, 268-272, 311-344/Disulfide bonds: #status predicted  
F.130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 13.4%; Score 289.5; DB 1; Length 388;  
Best Local Similarity 27.0%; Pred. No. 1.8e-16;  
Matches 103; Conservative 66; Mismatches 141; Indels 71; Gaps 16;

QY	3	DEEPEEGRGSGSEVMVDNLRKSGGCGYVEMTVGGPPQTLNLTVDTGSSNFAVGAAPH	62
		:    :    :    :    :    :    :    :    :    :    :    :    :    :	
Db	64	DEQPLE-----NYLDM-----EYFCIGICGTPAQDFTVFDTGSSNLNWPVSYCS	108
QY	63	FL-----HRYQROLSTSYDLRKGVVYPYTGKWEGLGTLVSIPIHGPNNVTVRANIAAI	118
		:    :    :    :    :    :    :    :    :    :    :    :    :    :	
Db	109	SLACTHNHRENPODSSTYQSTCTGVITTYGTSMTGILGYDTVQV--GGISDTNFIQGL	165
QY	119	TESDK-FFINGSNWEGILGLAYAEIARPDSDLSPFFPDSLVKQTHV-PNLFSLQCGAGPP	176
		:    :    :    :    :    :    :    :    :    :    :    :    :    :	
Db	166	SETPEGSFYAEPDFGILGLAYPSIS--SSGATPVEDTNWQGLSVQDILFVYLSAD---	220

[illegible]

Search completed: August 7, 2002, 09:14:18  
Job time: 42 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:15:11 ; Search time 17.62 Seconds  
(without alignments)  
894.373 Million cell updates/sec

Title: US-09-724-571-58  
Perfect score: 2156  
Sequence: 1 ETDEPEPEGRGSRFVEMV.....GPFVTLDMEDCGYNIPQTD 407

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2156	100.0	501	1	BACE_HUMAN
2	2134	99.0	501	1	BACE_RAT
3	2132	98.9	501	1	BACE_MOUSE
4	1132	52.5	518	1	BRE2_HUMAN
5	327	15.2	324	1	PEP1_GADMO
6	314.5	14.6	390	1	CATD_BOVIN
7	306	14.2	387	1	PEP1_RABIT
8	302.5	14.0	367	1	PEP4_MACFU
9	302	14.0	367	1	CATE_HUMAN
10	299.5	13.9	396	1	PEPE_CHICK
11	298.5	13.8	383	1	PEPE_CHICK
12	296	13.7	387	1	PEP2_RABIT
13	296	13.7	391	1	CATE_CAVPO
14	295.5	13.7	412	1	CATD_HUMAN
15	295	13.7	387	1	PEP4_RABIT
16	289.5	13.4	388	1	PEP2_MACFU
17	287	13.3	407	1	CATD_RAT
18	286	13.3	387	1	PEP3_RABIT
19	285.5	13.2	398	1	CATE_RAT
20	285	13.2	398	1	CATD_CHICK
21	283.5	13.1	388	1	PEPA_MACMU
22	282.5	13.1	388	1	PEPA_HUMAN
23	281.5	13.1	381	1	CHYM_SHEEP
24	281.5	13.1	388	1	PEP1_MOUSE
25	279.5	13.0	410	1	CATD_MOUSE
26	278.5	12.9	386	1	PEPA_PIG
27	278	12.9	387	1	PEPA_CALJA
28	277.5	12.9	396	1	CATE_RABIT
29	277.5	12.9	397	1	CATE_MOUSE
30	276.5	12.8	419	1	CHYM_CANAL
31	273.5	12.7	381	1	CHYM_BOVIN
32	273	12.7	388	1	PEP1_RABIT
33	270.5	12.5	377	1	PEP2_MACFU

34	270	12.5	376	1	PAG2_BOVIN
35	267.5	12.4	381	1	CHYM_CALJA
36	267	12.4	365	1	CATD_SHEEP
37	266	12.3	405	1	CARP_YEAST
38	264	12.2	396	1	CARP_NEUCR
39	263.5	12.2	388	1	PEPC_CALJA
40	262.5	12.2	394	1	PEPC_CAVPO
41	261.5	12.1	387	1	ASPP_AEDA
42	261.5	12.1	388	1	PEPC_HUMAN
43	261	12.1	388	1	PAG_HORSE
44	260	12.1	345	1	CATD_PIG
45	255	11.8	496	1	ASPR_ORYSA

## ALIGNMENTS

RESULT 1					
BACE_HUMAN					
ID	BACE_HUMAN	STANDARD	PRT	501 AA.	
AC	P56817; Q9UJT5;				
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)				
DE	(Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl protease 2) (Asp 2) (ASP2) (Membrane-associated aspartic protease 2) (Memapsin-2).				
DE	BACE OR BACE1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORM A).				
RC	TISSUE=Brain;				
RX	MEDLINE=20002972; PubMed=10531052;				
RA	Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A., Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y., Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A., Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F., Treanor J., Rogers G., Citron M.;				
RA	"Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE.";				
RT	Science 286:735-741(1999).				
RL					
RN					
RP	SEQUENCE FROM N.A. (ISOFORM A), SEQUENCE OF 46-68, AND CHARACTERIZATION.				
RC	TISSUE=Brain;				
RX	MEDLINE=20057171; PubMed=10591214;				
RA	Sinha S., Anderson J.P., Barbour R., Basl G.S., Caccavello R., Davis D., Doan M., Dovey H.F., Frigon N., Hong J., Jacobson-Croak K., Jewett N., Keim P., Knops J., Lieberburg I., Power M., Tan H., Tatsuno G., Tung J., Schenk D., Seubert P., Suenensaar S.M., Wang S., Walker D., Zhao J., McConlogue L., Varghese J.;				
RA	"Purification and cloning of amyloid precursor protein beta-secretase from human brain.";				
RT	Nature 402:537-540(1999).				
RL					
RN					
RP	SEQUENCE FROM N.A. (ISOFORM A).				
RC	MEDLINE=20057170; PubMed=10591213;				
RX	Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M., Brashier J.R., Straman N.C., Mathews W.R., Buhl A.E., Carter D.B., Tomasselli A.G., Parodi L.A., Heinrikson R.L., Gurney M.E.;				
RA	"Membrane-anchored aspartyl protease with Alzheimer's disease beta-secretase activity.";				
RT	Nature 402:533-537(1999).				
RL					
RN					
RP	SEQUENCE FROM N.A. (ISOFORM A).				
RC	MEDLINE=20030166; PubMed=10561122;				
RX	Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D., Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M., Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;				

Q28057	bos taurus
Q9n2d2	callithrix
Q9mzs8	ovis aries
P07267	saccharomyc
Q01294	neurospora
Q9n2d3	callithrix
Q64411	cavia porce
Q03168	aces aegyp
P20142	homo sapien
Q28389	equus cabal
P00795	sus scrofa
P42211	oryza sativ

"Identification of a novel aspartic proteinase (Asp 2) as beta-secretase.";  
Mol. Cell. Neurosci. 14:419-427(1999).

[5] SEQUENCE FROM N.A. (ISOFORM B).  
TISSEU-Brain, and Pancreas;  
Michel B., De Pietri Tonelli D., Zaccchetti D., Keller P.;  
"New beta-site APP cleaving enzyme isoform (BACE-1b) obtained from human brain and pancreas.";  
Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

[6] SEQUENCE OF 14-501 FROM N.A. (ISOFORM A), AND CHARACTERIZATION.

MEDLINE-20144060; PubMed-10677483;  
Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;  
"Human aspartic protease memapsin 2 cleaves the beta-secretase site of beta-amyloid precursor protein.";  
Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).

-!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP, LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL FRAGMENT WHICH IS LATER RELEASED BY GAMMA-SECRETASE.

-!- SUBCELLULAR LOCATION: Type I membrane protein.

-!- ALTERNATIVE PRODUCTS: 2 isoforms; A/BACE-1A (shown here) and B/BACE-1B; are produced by alternative splicing.

-!- TISSUE SPECIFICITY: BRAIN.

-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; AF190725; AAF04142.1; -  
DR EMBL; AF201466; AAF18982.1; -  
DR EMBL; AF200343; AAF17079.1; -  
DR EMBL; AF204943; AAF26367.1; -  
DR EMBL; AF338816; AAK38374.1; -  
DR EMBL; AF200193; AAF13715.1; -  
DR HSSP; P32329; 1YPS.  
MIM; 604252; -  
DR InterPro; IPR001969; Asp.protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF00026; asp; 3.  
DR PRINTS; PR00792; PEPSIN.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
KW Hydrolase; Aspartyl protease; Glycoprotein; zymogen; Transmembrane;  
FT SIGNAL; 1 21  
FT SIGNAL; Alternative splicing. POTENTIAL.  
FT PROPEP; 22 45  
FT CHAIN; 46 501  
FT DOMAIN; 22 457  
FT TRANSMEM; 458 478  
FT DOMAIN; 479 501  
FT ACT\_SITE; 93 93  
FT ACT\_SITE; 289 289  
FT ACT\_SITE; 153 153  
FT CARBOHYD; 172 172  
FT CARBOHYD; 223 223  
FT CARBOHYD; 354 354  
FT CARBOHYD; 190 214  
FT VARSPLIC; 501 AA; 55763 MW; 377CE4C824ACEF05 CRC64;  
SEQUENCE

Query Match 100.0%; Score 2156; DB 1; Length 501;  
Best Local Similarity 100.0%; Pred. No. 1.3e-171;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEPEGRGSGFVMDNLRGKSGQGYVEMTVGSPPTLNILNVDLTGSSNFAVGAAP 60  
Db 46 ETDEPEPEGRGSGFVMDNLRGKSGQGYVEMTVGSPPTLNILNVDLTGSSNFAVGAAP 105  
QY 61 HPFLHRYQROLSTYRDLRGVYVPYTOGKWEGLGELGTLVSIPIHGPNTVVRANIAITE 120  
Db 106 HPFLHRYQROLSTYRDLRGVYVPYTOGKWEGLGELGTLVSIPIHGPNTVVRANIAITE 165  
QY 121 SDKFFINGSNNEGILGLAYAEIARPDSDLEFPFDSLVKQTHVPLNLSIQLCGAGFPLNQ 180  
Db 166 SDKFFINGSNNEGILGLAYAEIARPDSDLEFPFDSLVKQTHVPLNLSIQLCGAGFPLNQ 225  
QY 181 EVLASVGSGMIIGGDHSLYTGSLWYTPIRREWYEVIIIRVEINGQDLKMDCKEYNDK 240  
Db 226 EVLASVGSGMIIGGDHSLYTGSLWYTPIRREWYEVIIIRVEINGQDLKMDCKEYNDK 285  
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASTKFPDGFGLGOLVCWQAGTTPWNIFPVIS 300  
Db 286 SIVDSGTTNLRPKKVFEEAAVKSIAASTKFPDGFGLGOLVCWQAGTTPWNIFPVIS 345  
QY 301 LYLMGEVNTQSRITLPOQYLPRVEDVATSDQDCYKFAISQSSGTVMGAVIMEGFYV 360  
Db 346 LYLMGEVNTQSRITLPOQYLPRVEDVATSDQDCYKFAISQSSGTVMGAVIMEGFYV 405  
QY 361 FDRARRKRGFAYSACHVHDEFRTAAVEGPFVTLDMDCGYNIPTQDE 407  
Db 406 FDRARRKRGFAYSACHVHDEFRTAAVEGPFVTLDMDCGYNIPTQDE 452

RESULT 2

ID	BACE-RAT	STANDARD	PRT	501 AA.
AC	P56819;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Beta-secretase precursor (EC 3.4.23.-) (beta-site APP cleaving enzyme)			
DE	Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl			
DE	protease 2) (Asp 2) (ASP2) (Membrane-associated aspartic protease 2)			
DE	(Memapsin-2).			
GN	BACE.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=101116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20002972; PubMed=10531052;			
RA	Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,			
RA	Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,			
RA	Fisher S., Fuller J., Edenson S., Lille J., Jarosinski M.A.,			
RA	Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,			
RA	Treanor J., Rogers G., Citron M.			
RT	"Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE.";			
RL	Science 286:735-741(1999).			
CC	-!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP, LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL FRAGMENT WHICH IS LATER RELEASED BY GAMMA-SECRETASE (BY SIMILARITY).			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ). -----			



Best Local Similarity 98.3%; Pred. No. 1.3e-169;		Matches 400; Conservative 4; Mismatches 3; Indels 0; Gaps 0;	
Qy	1	ETDEPEPEGRGSGFVEMVNLKSGQGYVEMTVGSPPTLNILVDTGSSNFVAGAAP	60
Db	46	ETDESEEPGRGSGFVEMVNLKSGQGYVEMTVGSPPTLNILVDTGSSNFVAGAAP	105
Qy	61	HPFLHRYQRLSTYSYRDLKRGVVPYVYQGWEGELGTDLVSPHGNVTVRANIAITE	120
Db	106	HPFLHRYQRLSTYSYRDLKRGVVPYVYQGWEGELGTDLVSPHGNVTVRANIAITE	165
Qy	121	SKDFFINGNWEGILGLAYAEIAPDDSLVFKQTHVNLFLSLQLCGAGFPLNQS	180
Db	166	SKDFFINGNWEGILGLAYAEIAPDDSLVFKQTHVNLFLSLQLCGAGFPLNQT	225
Qy	181	EVLASVGSMIIGGIDHSLYTGSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNDK	240
Db	226	EALASVGSMIIGGIDHSLYTGSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNDK	285
Qy	241	SIVDSGTTNLRPKVFAAVKSKAASSTKFPDGFGLGQVLCWAGTTPWNIFPVIS	300
Db	286	SIVDSGTTNLRPKVFAAVKSKAASSTKFPDGFGLGQVLCWAGTTPWNIFPVIS	345
Qy	301	LYLMGEVNTQSFRTILPQQYLRPVEDVATSDDCYKFAISQSTGTVMGAVIMEGFVV	360
Db	346	LYLMGEVNTQSFRTILPQQYLRPVEDVATSDDCYKFAISQSTGTVMGAVIMEGFVV	405
Qy	361	FDRARKRIGFAVSACHVHDEFRTAAVEGFFVTLDMEDCGYNIPQDTE	407
Db	406	FDRARKRIGFAVSACHVHDEFRTAAVEGFFVTLDMEDCGYNIPQDTE	452
RESULT 4			
ID	BAE2_HUMAN	STANDARD:	PRT; 518 AA.
AC	Q9V5Z0; Q9UJTG;		
DC	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DE	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Beta secretase 2 precursor (EC 3.4.23.-) (Beta-site APP-cleaving		
DE	enzyme 2) (Aspartyl protease 1) (Asp 1) (ASPI) (Membrane-associated		
DE	aspartic protease 1) (Memapsin-1).		
GN	BACE2 OR ASP21.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE=20057170; PubMed=10591213;		
RA	Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,		
RA	Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,		
RA	Tomasselli A.G., Parodi L.A., Heinrichson R.L., Gurney M.E.;		
RT	"Membrane-anchored aspartyl protease with Alzheimer's disease		
RT	beta-secretase activity.";		
RL	Nature 402:533-537(1999).		
RN	[2]		
RN	SEQUENCE FROM N.A.		
RC	TISSUE=Bone marrow;		
RA	Xin H., Stephens J.C., Duan X., Harrowe G., Kim E., Grieshammer U.,		
RA	Giese K.;		
RT	"Identification of a novel aspartic-like protease differentially		
RT	expressed in human breast cancer cell lines.";		
RT	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RN	SEQUENCE FROM N.A.		
RA	Accarino M.P., Funagalli P., Ottolenghi S., Taramelli R.;		
RT	"Cloning of a gene from chromosome 21 Down region encoding a potential		
RT	transmembrane aspartyl protease.";		
RT	Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.		
RN	[4]		
RN	SEQUENCE FROM N.A.		
RA	Solans A., Estivill X., de la Luna S.;		

RT	"Cloning of a novel mammalian aspartyl protease.";	
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.	
RN	[5]	
RN	SEQUENCE FROM N.A.	
RX	MEDLINE=20030166; PubMed=10561122;	
RA	Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,	
RA	Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M.,	
RA	Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;	
RT	"Identification of a novel aspartic proteinase (Asp 2) as	
RT	beta-secretase.";	
RL	Mol. Cell. Neurosci. 14:419-427(1999).	
RN	[6]	
RN	SEQUENCE FROM N.A.	
RX	MEDLINE=20144060; PubMed=10677483;	
RA	Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;	
RT	"Human aspartic protease memapsin 2 cleaves the beta-secretase site of	
RT	beta-amyloid precursor protein.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).	
RN	[7]	
RN	SEQUENCE FROM N.A.	
RX	MEDLINE=20289799; PubMed=10830953;	
RA	Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,	
RA	Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,	
RA	Onki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,	
RA	Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,	
RA	Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,	
RA	Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Antonarakis S.E.,	
RA	Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Brandt P.,	
RA	Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Kauer G., Bloeker H.,	
RA	Scharfe M., Schoen O., Desario A., Reichelt J., Rieselmann L., Dagand E.,	
RA	Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,	
RA	Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,	
RA	Lehrach H., Reinhardt R., Yaspo M.-L.;	
RT	"The DNA sequence of human chromosome 21.";	
RL	Nature 405:311-319(2000).	
CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.	
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE	
CC	EUKARYOTIC ASPARTYL PROTEASES FAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL; AF200342; AAF17078.1; -	
DR	EMBL; AF117892; AAD45240.1; -	
DR	EMBL; AF050171; AAD45963.1; -	
DR	EMBL; AF178532; AAF29494.1; -	
DR	EMBL; AF204944; AAF26368.1; -	
DR	EMBL; AF200192; AAF13714.1; -	
DR	EMBL; AL163284; CAB90458.1; -	
DR	EMBL; AL163285; CAB90554.1; -	
DR	HSP; P00797; 2REN.	
DR	MM; 605668; -	
DR	InterPro; IPR001969; Asp-protease.	
DR	InterPro; IPR001461; Pepsin.	
DR	Pfam; PF00026; asp; 3.	
DR	PRINTS; PR00792; PEPSIN.	
DR	PROSITE; PS00141; ASP-PROTEASE; 2.	
KW	Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;	
KW	Signal.	
FT	SIGNAL	1 20
FT	PROPEP	21 ?
FT	CHAIN	? 518
FT	DOMAIN	21 473
FT	TRANSMEM	474 494
FT	DOMAIN	495 518
FT	ACT_SITE	110 110
FT	ACT_SITE	303 303
FT	CARBOHYD	170 170
FT	POTENTIAL.	
FT	BETA SECRETASE 2.	
FT	EXTRACELLULAR (POTENTIAL).	
FT	POTENTIAL.	
FT	CYTOPLASMIC (POTENTIAL).	
FT	BY SIMILARITY.	
FT	BY SIMILARITY.	
FT	N-LINKED (GLCNAC... ) (POTENTIAL).	





CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
DR PIR: S32383; S32383.  
DR PIR: S37419; S37419.  
DR HSP: P07339; 1LYB.  
DR MEROPS: A01.009; .  
DR InterPro: IPR001969; Asp\_protease.  
DR InterPro: IPR001461; Pepsin.  
DR Pfam: PF00026; asp; 1.  
DR PRINTS: PR00792; PEPsin.  
DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
KW Hydrolyase: Aspartyl protease; Glycoprotein; Lysosome; Zymogen.  
FT PROPEP 1 44  
FT CHAIN 45 390  
FT ACT\_SITE 77 77  
FT ACT\_SITE 273 273  
FT DISULFID 71 140  
FT DISULFID 90 97  
FT DISULFID 264 268  
FT DISULFID 307 344  
FT CARBOHYD 114 114  
FT CARBOHYD 241 241  
SQ SEQUENCE 390 AA; 42488 MW; 5B38AA1C33C48D35 CRC64;  
  
Query Match 14.6%; Score 314.5; DB 1; Length 390;  
Best Local Similarity 28.0%; Pred. No. 1e-18;  
Matches 113; Conservative 72; Mismatches 128; Indels 91; Gaps 21;  
  
QY 8 EPG-RRGSFVEMVDNLKSGQGYVEMTGGSPPTNLNLDVDTGSSNFAVGAAPFL-- 64  
DB 39 EFAVRGPIPELLKNYMDAQ---YYGIGIGTPPQCFTVFDGTSANLWVPSIHCKLLDI 95  
QY 65 ----HRYQRQLSSYVRDLRKGVY-VPTQCKWEGELGTLVSTPHGPN-----VTVR 112  
DB 96 ACWTHRKYSKSSY--VKNGTTFDIHYGSGSLGYSQDTSVPCNPSSSPGCVIVQ 153  
QY 113 ANI--AAITSDKFFNGWEGILGLAYAEIARPDLSLEPFDLSLVKQTHV-PNLFSIQ 169  
DB 154 RQTGFAIKPGVVF--AAKFDGILGMAYPRIS--VNNVLPVFDNLQKLVKDNVFS-- 208  
QY 170 LCGAGFPLNOSVLAAGSGMIIGDHSYTGSLWYTPIREWYEVIVRVEINGQDL 229  
DB 209 -----FPLNR-DPKAQGGELMLGGTDSKYRGSLMFHNTVQVQIIMDQDLV-GSSL 261  
QY 230 KMDCKEYNDKSTVDSTTNLRPKVFEAAVKAASSTKFPDGFVLGEOLV-CWQA 288  
DB 262 TV-CK--GGCEALVDGTSLVGPVEVRELQKAGVPLIQ-----GEVMPCEKV 310  
QY 289 GTTPWNIFPVISLYLMGEVNTQSFRTIILPQQYLRVEDVATSDQDCYKFAISQSTGT- 347  
DB 311 SS-----LPEVTVKLG-----KDYALSPED-YALKVSAETTVVC 344  
QY 348 -----VMGAVIMEGFYVVDVDRKRIGFAVSA 374  
DB 345 LSGFMGMDIPPPGGLMILGDVFIIGRYITVFDQNRVGLAEAA 388  
  
RESULT 7  
ID PEPI\_RABIT STANDARD; PRT; 387 AA.  
AC P28712;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Pepsin II-1 precursor (EC 3.4.23.1) (Pepsin A).  
OS Eryctolagus cuniculus (Rabbit)  
OC Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Eukaryota; Metazoa; Lagomorpha; Leporidae; Oryctolagus.  
OC NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91009127; PubMed=2129536;

RA Kageyama T., Tanabe K., Koizumi O.;  
RT "Structure and development of rabbit pepsinogens. Stage-specific  
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and  
RT gene expression during development.";  
RL J. Biol. Chem. 265:17031-17038(1990).  
CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS  
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE  
CC ALSO CLEAVED TO SOME EXTENT.  
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.  
CC -1- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE  
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.  
CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY  
CC HORMONES AND RELATED SUBSTANCES.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
DR PIR: B38302; B38302.  
DR HSP: P00791; 1PSA.  
DR MEROPS: A01.001; .  
DR InterPro: IPR001969; Asp\_protease.  
DR InterPro: IPR001461; Pepsin.  
DR Pfam: PF00026; asp; 1.  
DR PRINTS: PR00792; PEPsin.  
DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
KW Hydrolyase: Aspartyl protease; Digestion; Stomach; Gastric juice;  
KW Zymogen; Signal; Phosphorylation; Multigene family.  
FT SIGNAL 1 15  
FT PROPEP 16 59  
FT CHAIN 60 387  
FT MOD\_RES 129 129  
FT ACT\_SITE 93 93  
FT ACT\_SITE 276 276  
FT DISULFID 106 111  
FT DISULFID 267 271  
FT DISULFID 310 343  
FT SEQUENCE 387 AA; 42070 MW; A6EC48F715541A48 CRC64;  
  
Query Match 14.2%; Score 306; DB 1; Length 387;  
Best Local Similarity 27.1%; Pred. No. 5.2e-18;  
Matches 98; Conservative 67; Mismatches 131; Indels 66; Gaps 15;  
  
QY 30 YVEMTVGSPPTNLNLDVDTGSSNFAVG----AAPHPLHRYQQLSTYDLRKGVYV 85  
DB 75 YFGTISIGTPPQEFVIEDTSSNLWVPESTYCSLACFLHRRFDDSDSTFQATSETLSI 134  
QY 86 PYTQCKWEGELGTLVSTPHGPNVTVRANIAAITESED---KFFINGSNWEGILGLAYAEI 142  
DB 135 TYGTGSMTGILGYDITKV---GNIEDTNQIFGLSKTEPGITFLV--APFDGILGLAYPSI 189  
QY 143 ARPDLSLEPFDLSLVKQTHV-PNLFSLQCGAGFPLNOSVLAAGSGMIIGDHSLYT 201  
DB 190 SASDAT--PVFDNMWNEGLVSEDLFSVYLSNG-----ERKSMVMEFGDSSYTT 237  
QY 202 GSWTTPPIRREWYEVIVRVEINGQDLKM--DCKEYNDKSTVDSTTNLRPKVFEA 259  
DB 238 GSWTTPPIRREWYEVIVRVEINGQDLKM--DCKEYNDKSTVDSTTNLRPKVFEA 291  
QY 260 AVKSIKAASSTKFPDGFVLGEOLV-CWQAGTTPWNIFPVISLYLMGEVNTQSFRTIILP 318  
DB 292 IQSYIGASKNL-----LGENIISCAISDLPDIVE-----TINN 325  
QY 319 QQYLREVED-VATSDQDC---YKFAISQSTGT--VMGAVIMEGFYVVDVDRKRIGFAV 372  
DB 326 VQYPLPASAYILKEDDDCLSGFDGMNLDTSYGLWILGDVFIIRQYVTFVDRANNQVGLAA 385  
QY 373 SA 374  
DB 386 AA 387  
  
RESULT 8  
ID PEP4\_MACFU STANDARD; PRT; 388 AA.  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91009127; PubMed=2129536;

P27678;  
 01-AUG-1992 (Rel. 23, Created)  
 01-AUG-1992 (Rel. 23, Last sequence update)  
 01-NOV-1995 (Rel. 32, Last annotation update)  
 Pepsin A-4 precursor (EC 3.4.23.1) (Pepsin I/II).  
 PGA.  
 Macaca fuscata fuscata (Japanese macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
 OC Cercopitheinae; Macaca.  
 NCBI\_TaxID=9543;  
 [1]  
 SEQUENCE FROM N.A., AND SEQUENCE OF 16-70.  
 RC TISSUE=Gastric mucosa;  
 RX MEDLINE=92037645; PubMed=1935977;  
 RA Kageyama T., Tanabe K., Koiwai O.;  
 RT "Developments-dependent expression of isozymogens of monkey  
 RT pepsinogens and structural differences between them.";  
 RL Eur. J. Biochem. 202:205-215(1991).  
 CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS  
 CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE  
 CC ALSO CLEAVED TO SOME EXTENT.  
 CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.  
 CC -!- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY  
 CC HORMONES AND RELATED SUBSTANCES.  
 CC -!- MISCELLANEOUS: EACH PEPSINOGEN IS CONVERTED TO CORRESPONDING  
 CC PEPSIN AT PH 2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA  
 CC ACTIVATION SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC  
 CC CLEAVAGE VIA AN INTERMEDIATE FORM(S).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X59753; CAA42425.1; -;  
 DR PIR; S16065; S16065.  
 DR PIR; S19682; S19682.  
 DR HSP; P00790; IPSN.  
 DR MEROPS; A01.001; -;  
 DR InterPro; IPR001969; Asp.protease.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp; 1.  
 DR PRINTS; PR00792; PEPsin.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 2.  
 DR Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;  
 KW Zymogen; Multigene family; Signal; Glycoprotein.  
 FT SIGNAL 1 15  
 FT PROPEP 16 38  
 FT PROPEP 39 62  
 FT CHAIN 63 388  
 FT ACT\_SITE 94 94  
 FT ACT\_SITE 277 277  
 FT DISULFID 107 112  
 FT DISULFID 268 272  
 FT DISULFID 311 344  
 FT CARBOHYD 88 88  
 SEQUENCE 388 AA; A1955 MW; A2923AB1F7FCDEB9 CRC64;  
 Query Match 14.0%; Score 302.5; DB 1; Length 388;  
 Best Local Similarity 27.6%; Pred. No. 1e-17;  
 Matches 107; Conservative 64; Mismatches 133; Indels 83; Gaps 17;  
 QY 3 DEEPEPGRGSEVMVDNLRGSGGQYVEMVGPPTNLNLTVDGSSNFAVGAAPHP 62  
 DB 64 DEOPLE-----NVLDV-----EYFGTIGTIGTIPAQNFVFTGSSNLWV----PSV 105

QY 63 FL-----HRYQRLSSYRDLRKGVVVPTQCKWEGELGTDLVSPGPNVTVRANI 115  
 DB 106 YCYSLACMDHNLFPQDSSTYRATSKTVSYTYGTGSMTGILGYDTRKV---GGISDTNQI 162  
 QY 116 AAITESDK-FFTINGSNWEGILGAYAEIARPDSDLSEPPFDSLVRKQTHV-PNLFSLQLCGA 173  
 DB 163 FGLSETEPGFFLYFAPDFGILGLAYPSIS--SSGATPVFDNIWQRLVSQDLFSVYLSAD 220  
 QY 174 GPPLNQSEVLASVGGMIIGGIDHSLYTGSLWYTPIRREWYEVYIIVRVEINGQDL--KM 231  
 DB 221 ----DQS-----GSVIFGGIDSSYTGSLNWVPVSVEGYWQISVDSITMNGKTIACAK 270  
 QY 232 DKEYNYDKSIVDSGTTNLRPKKVFEEAAVKSIIKAASSTKFPDGFWLGEOLV-CWQAGT 290  
 DB 271 GC-----QAIVDTGTSLLTGTPSIANIQSDIGASENSD-----GENVYSCSAISS 316  
 QY 291 TPNWIFPVISLYLMEVYNQSFRTILPQY-LRPVEDVATSDQCYK-----FAISQSS 344  
 DB 317 LPDIVF-----TINGVQYPLPPSAYILQSQGSCSTSGFGMDVPTEG 358  
 QY 345 TGTWGAIVMEGFYVDFDRARKRIGFA 371  
 DB 359 ELWILGDVFIQYVTFVDRANNQVGLA 385  
 RESULT 9  
 PEPA\_CHICK  
 ID PEPA\_CHICK STANDARD; PRT; 367 AA.  
 AC P00793;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DE 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Pepsin A precursor (EC 3.4.23.1).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=84004412; PubMed=6617663;  
 RA Baudys M., Kostka V.;  
 RT "Covalent structure of chicken pepsinogen.";  
 RL Eur. J. Biochem. 136:89-99(1983).  
 CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS  
 CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE  
 CC ALSO CLEAVED TO SOME EXTENT.  
 CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 DR PIR; A00984; PECH.  
 DR HSP; P00794; 4CMS.  
 DR MEROPS; A01.001; -;  
 DR InterPro; IPR001969; Asp.protease.  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp; 1.  
 DR PRINTS; PR00792; PEPsin.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 2.  
 KW Hydrolase; Aspartyl protease; Digestion; Stomach; Zymogen;  
 KW Glycoprotein; Gastric juice.  
 FT PROPEP 1 42  
 FT CHAIN 43 367  
 FT ACT\_SITE 77 77  
 FT ACT\_SITE 260 260  
 FT CARBOHYD 113 113  
 FT DISULFID 90 95  
 FT DISULFID 251 255  
 FT DISULFID 290 323  
 SQ SEQUENCE 367 AA; 40431 MW; 0C547E7ED8F5B341 CRC64;  
 Query Match 14.0%; Score 302; DB 1; Length 367;

30	Qy	YVEMVTGSPPOTLNLLVDTGSSNFAVCAAPHPL-----HRYTORLSSTYRLDKRG	82
		:    :    :    :    :    :    :    :    :	
59	Db	YFGPISITGPOQDFSVIFDTGSSNLWV---PSIYCKSSACSNHKREDPSKSTVYSTNET	115
		:    :    :    :    :    :    :    :    :	
83	Qy	VVPYTOCKWEGELGDLVSIPIGPNVTVRANIAAITESDK -FFINGSNWEGILGLAYAE	141
		:    :    :    :    :    :    :    :    :	
116	Db	VYIAYGTGSMGILGYDTAV---SSIDVQNOIFGLSETEPGFFYVCNFDGILGLAPPS	172
		:    :    :    :    :    :    :    :    :	
142	Qy	IARDDDSLEPPFDLSLVKQTHV -PNLFSLQLCGAGFPLNQSEVLASVGGSMILGGIDHSLY	200
		:    :    :    :    :    :    :    :    :	
173	Db	IS--SSGATPVFDNMMSQHLVAQDLFSVLSKDG-----ETGSFVLFGGIDPNVT	220
		:    :    :    :    :    :    :    :    :	
201	Qy	TGSLWYPIRREWYEVILVRIEINGODLK -MCKEYNDKSVISGTTNLRPKKVE	258
		:    :    :    :    :    :    :    :    :	
221	Db	TGKIYVWPLSAREYQWITMDRVTVGNKYVACFFTC-----QAIVDGTSLVWPOGAYN	274
		:    :    :    :    :    :    :    :    :	
259	Qy	AAVKSIRKAASSTE-----KFPDGFWLGEQLVCMQAGTTPNTPFVISLYLMGEVTNOS	311
		:    :    :    :    :    :    :    :    :	
275	Db	RIIKDLGVSSDGEISCDDISKLPD-----VTFPHNGHA-----	307
		:    :    :    :    :    :    :    :    :	
312	Qy	FRITILPOQYLRPVEDVATSQDDCYKFAISQSTGT-----VMGAVIMEGFYVVDRAR	365
		:    :    :    :    :    :    :    :    :	
308	Db	-----FTLPASAYVLNEDGSCMLGFENMGTPTELGEQWILGDVFIREYVIFEDRAN	358
		:    :    :    :    :    :    :    :    :	
366	Qy	KRIGFA	371
		:	
359	Db	NRVGLS	364
		:	

RESULT	10				
CATE_HUMAN					
ID	CATE_HUMAN	STANDARD;	PRT;	396	AA.
AC	P14091;				
DT	01-JAN-1990	(Rel. 13, Created)			
DT	01-JAN-1990	(Rel. 13, Last sequence update)			
DT	01-MAR-2002	(Rel. 41, Last annotation update)			
DE	Cathepsin E precursor	(EC 3.4.23.34).			
GN	CTSE				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa;	Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxId:9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=89380302; PubMed=2674141;				
RA	Azuma T., Pals G., Mohandas T.K., Couvreur J.M., Taggart R.T.;				
RA	"Human gastric cathepsin E. Predicted sequence, localization to				
RT	chromosome 1, and sequence homology with other aspartic				
RT	proteinases.";				
RT	J. Biol. Chem. 264:16748-16753(1989).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=92112877; PubMed=1370478;				
RA	Azuma T., Liu W.G., Vander Laan D.J., Bowcock A.M., Taggart R.T.;				
RA	"Human gastric cathepsin E gene. Multiple transcripts result from				
RT	alternative polyadenylation of the primary transcripts of a single				
RT	gene locus at 1q31-q32.";				
RT	J. Biol. Chem. 267:1609-1614(1992).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RA	Tatnell P.J., Kay J.;				
RA	"Human procathepsin E.";				
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.				
RN	[4]				
RP	SEQUENCE OF 54-68: 77-95; 141-155; 275-285 And 389-396.				
RX	MEDLINE=90241267; PubMed=2334440;				
RA	Athauda S.B.P., Matsuzaki O., Kageyama T., Takahashi K.;				
RT	"Structural evidence for two isozymic forms and the carbohydrate				
RT	attachment site of human gastric cathepsin E.";				
RT	Biochem. Biophys. Res. Commun. 168:878-885(1990).				

!- FUNCTION: DUE OT ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN  
LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.  
!- CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader  
specificity.  
!- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.  
!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
EUKARYOTIC ASPARYL PROTEASES FAMILY.

EMBL; M84424; AAA52300.1; -. JOINED.	18	53	ACTIVATION PEPTIDE.
EMBL; M84413; AAA52300.1; JOINED.	54	396	CATHEPSIN E. PYRROLIDONE CARBOXYLIC ACID.
EMBL; M84417; AAA52300.1; JOINED.	18	18	BY SIMILARITY.
EMBL; M84418; AAA52300.1; JOINED.	96	96	BY SIMILARITY.
EMBL; M84419; AAA52300.1; JOINED.	281	281	INTERCHAIN (PROBABLE).
EMBL; M84420; AAA52300.1; JOINED.	60	60	BY SIMILARITY.
EMBL; M84421; AAA52300.1; JOINED.	109	114	BY SIMILARITY.
EMBL; M84422; AAA52300.1; JOINED.	272	276	BY SIMILARITY.
EMBL; J05036; AAA52130.1; -. N-LINKED (GLCNAC. . .).	314	351	N-LINKED (POTENTIAL).
EMBL; AJ250717; CAB82850.1; -. O-LINKED (POTENTIAL).	90	90	O-LINKED (POTENTIAL).
EMBL; A34401; A34401.	220	220	333
PIR; A34643; A34643.	336	336	AA; 42793 MW; 40B643C5F801521E CRC64;
PIR; A42038; A42038.			
HSP; P00794; 4CMS.			
MEROFS; A01.010; -. MIM; 116890; -. InterPro; IPR001969; Asp_protease.			
InterPro; IPR001461; Pepsin.			
PIR; PF00026; asp; 1.			
PRINTS; PR00792; PEPSIN.			
PROSITE; PS00141; ASP_PROTEASE; 2.			
Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Signal.			
SIGNAL	1	17	
PROPEP	18	53	
CHAIN	54	396	
MOD_RES	18	18	
ACT_SITE	96	96	
ACT_SITE	281	281	
DISULFID	60	60	
DISULFID	109	114	
DISULFID	272	276	
DISULFID	314	351	
CARBOHYD	90	90	
CARBOHYD	220	220	
CARBOHYD	333	333	
SEQUENCE	336	336	
SO			

```

Query Match      13.98; Score 299.5; DB 1; Length 396;
Best Local Similarity 25.98; Pred. No. 1.9e-17;
Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;

      3 DEEPEEPCRRGSGFEMVDNLKSGQGYVYEMTVGSPPTLNILVDGSSNFVCA----58
      l: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
      63 DQSAKEP-----LNYLD-----MEYFGTISGSPQNFVFIDTGSNLLWPSVYCT 110
      l: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

      59 APHPFLHRYQORLSSYTRDLRKGYYVPTQGWEGELGTLVSLPHGPNVTVRANTAAI 118
      l: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

      111 SPACKTHSRFPQSSSTVSQPGSFSIQYTGSLSGIICADQSV-EGLTVVGQFGESV 169
      l: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

      119 TESDKFFINGSNWEGILGLAYAEIARPDSDSLEPFDDSLVKQTHVNPVLSQLCGAGFFPN 178
      l: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

      170 TEPGQTFVD-AEEDGILGLGYPSLA--VGGVTVPVDNNMAQ-----NLVDLPMSFYMSSN 222
      l: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

      179 QSEVLASVGSMMIIGDHSLYTGSWLYTPTRERWYVIVVRVEINGDLKMDCKEYNY 238
      l: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

      223 PE---GGAGSELIFGQYDHSFSGSLNWPVPTKQAYWQIALDNIQVGG--TWMFCSF--G 275

```



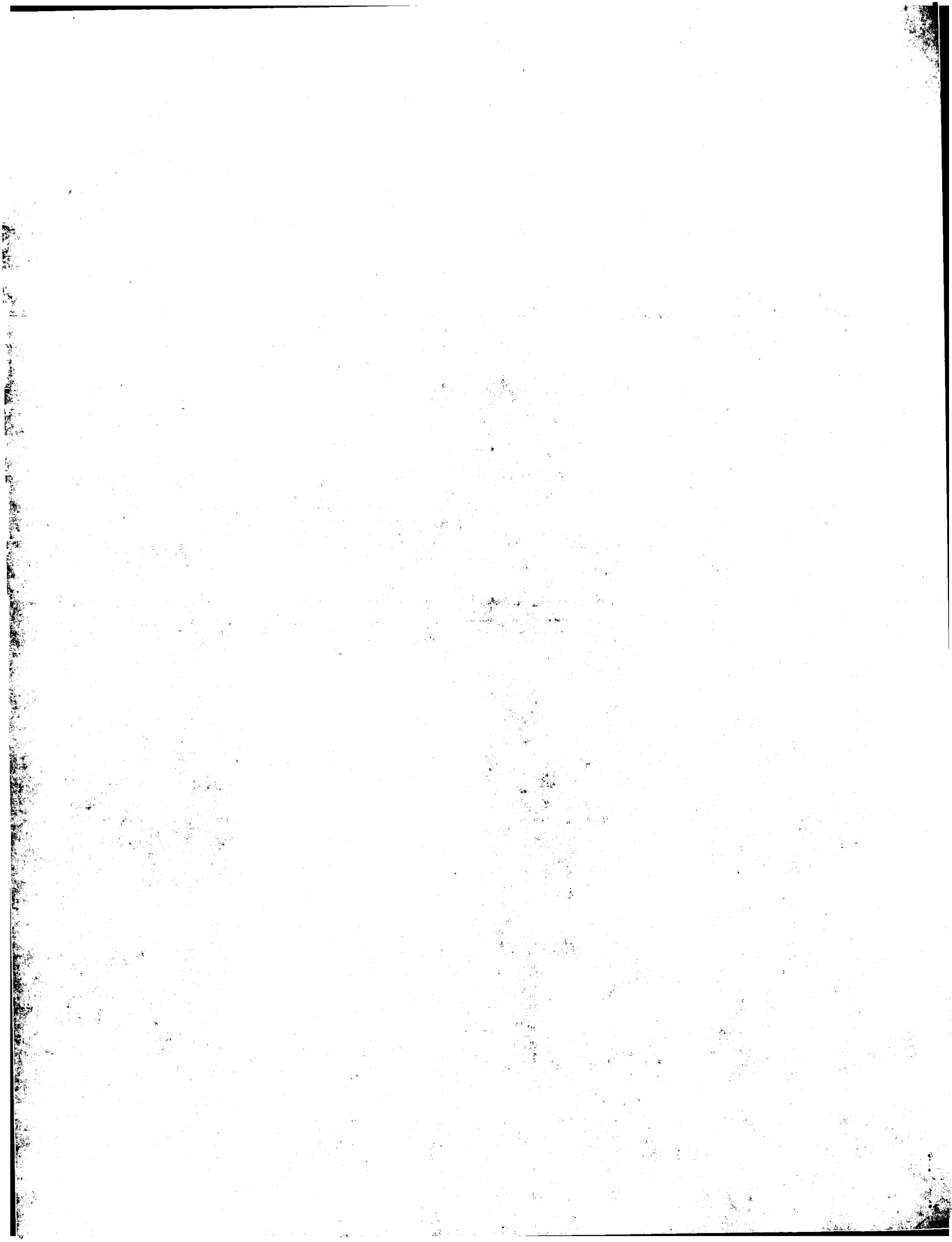
Pfam: PF00026; asp: 1.  
DR PROSITE; PRO0792; PEPsin.  
DR PROSITE; PS00141; ASP-PROTEASE; 2.  
DR Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;  
KW Zymogen; Signal; Phosphorylation; Multigene family.  
FT SIGNAL 1 15  
FT PROPEP 16 59  
FT CHAIN 60 387  
FT MOD\_RES 129 129  
FT ACT\_SITE 93 93  
FT ACT\_SITE 276 276  
FT ACT\_SITE 106 111  
FT DISULFID 267 271  
FT DISULFID 310 343  
FT SEQUENCE 387 AA; 42100 MW; 66FC331A3DC75891 CRG64;  
  
Query Match 13.7%; Score 296; DB 1; Length 387;  
Best Local Similarity 26.9%; Pred. No. 3.5e-17; Indels 66; Gaps 13;  
Matches 97; Conservative 63; Mismatches 135;  
  
QY 30 YVEMTVGSPPTLNILVDGTGSSNFAYGAAPHPF-----LHRYQQLSSYRDLRKG 82  
DB 75 YFTISIGTPPDFTVFDTGSSNLV---PSTYCSSLACALHKRNPEDSSTYQGTSET 131  
QY 83 VVVPYTOGKWEGLGTDLVSPHGNVTVRANIAATIESDKFFINGSNWEGILGLAYAEI 142  
DB 132 LSITYTGSMTGILGYDVKVGSIEDTQIFGLSKTEPSLTLF--APFDGILGLAYPSI 189  
QY 143 ARPDSLEPFDLSLVKQTHV--PNFLSLQCGAGPFLNQSEVLASVGGSMIIGGDHSLYT 201  
DB 190 SSDAT--PVFNMMNEGILVSQDLSEVYLSDD-----EKSILVMFGGIDSSVYT 237  
QY 202 GSWLTPPIRREYVEIIVRVEINGQDLKM--DCKEYNDKSIYDSTGNTNLRPKVFEA 259  
DB 238 GSWLWVPVSEYQWITMDSVINGETACADSC-----QAIVDTGSLTGP-----TS 287  
QY 260 AVKSIKAASSTKFPDGLGQLV--CQAGTTPNIPFPVSLVLMGEVNTQSFRTILP 318  
DB 288 AISNIQSYIGASK-----NLGENVISCSAIDSLDIVF-----TING 325  
QY 319 QQLRLRVEDVATSDQDCYKFAISQSSTGT-----VMGAVIMEGYVYVDFDRKRGFAV 372  
DB 326 IQYPLPASVILKEDDCTSGLEGMNVDYTGELMTLGDVFIQYFTVDFRANNOGLAA 385  
QY 373 S 373  
DB 386 A 386  
  
RESULT 13  
CATE\_CAVPO STANDARD; PRT; 391 AA.  
ID CATE\_CAVPO  
AC P25796;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Cathepsin E precursor (EC 3.4.23.34).  
GN Crse.  
OS Cavia porcellus (Guinea pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.  
OX NCBI\_TaxID=10141;  
RN [1]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RX MEDLINE=92355614; PubMed=1644829;  
RA Kageyama T., Ichinose M., Tsukada S., Miki K., Kurokawa K., Koizumi O.,  
Tajiri M., Yakabe E., Achauda S.B., Takahashi K.;  
RT "Gastric procathepsin E and procathepsin from guinea pig.  
RT Purification, molecular cloning of cDNAs, and characterization of  
RT enzymatic properties, with special reference to procathepsin E";  
RL J. Biol. Chem. 267:16450-16459(1992).  
RN [2]  
  
SEQUENCE FROM N.A.  
RC TISSUE=Gastric mucosa;  
RX MEDLINE=96073637; PubMed=8540321;  
RA Kageyama T., Ichinose M., Miki K., Moriyama A., Yonezawa S.,  
Tajiri M., Achauda S.B., Takahashi K.;  
RT "Isolation, characterization, and structure of procathepsin E and  
RT cathepsin E from the gastric mucosa of guinea pig";  
FT Signal 1 15  
FT PROPEP 16 59  
FT CHAIN 60 387  
FT MOD\_RES 129 129  
FT ACT\_SITE 93 93  
FT ACT\_SITE 276 276  
FT ACT\_SITE 106 111  
FT DISULFID 267 271  
FT DISULFID 310 343  
FT SEQUENCE 387 AA; 42100 MW; 66FC331A3DC75891 CRG64;  
  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
EMBL; M88653; AAA37052.1; -;  
EMBL; S80547; AAB35844.1; -;  
PIR; A43356; A43356.  
HSSP; P00794; 4CMS.  
MEROPS; A01.010; -;  
InterPro: IPR001969; Asp-protease.  
InterPro: IPR001461; Pepsin.  
Pfam: PF00026; asp: 1.  
PROSITE; PRO0792; PEPsin.  
PROSITE; PS00141; ASP-PROTEASE; 2.  
Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Signal.  
FT SIGNAL 1 18  
FT PROPEP 19 53  
FT CHAIN 54 391  
FT ACT\_SITE 92 92  
FT ACT\_SITE 276 276  
FT DISULFID 56 56  
FT DISULFID 105 110  
FT DISULFID 267 271  
FT DISULFID 309 346  
FT CARBOHYD 86  
FT SEQUENCE 391 AA; 42132 MW; 78D216B8FCDBD CRC64;  
  
Query Match 13.7%; Score 296; DB 1; Length 391;  
Best Local Similarity 26.9%; Pred. No. 3.6e-17; Indels 72; Gaps 16;  
Matches 98; Conservative 64; Mismatches 130;  
  
QY 30 YVEMTVGSPPTLNILVDGTGSSNFAYGA-----APHPFLHRYQQLSSYRDLRKGYYV 85  
DB 74 YFTISIGSPPTNFVFDTGSSNLVPSVYCTSPACQTHPVFHPSLSSYREVGNFSI 133  
QY 86 PYTOGKWEGLGTDLVSPHGNVTVRANIAATIESDKFFINGSNWEGILGLAYAEIAR 145  
DB 134 QXGTGSLTGIIGADQSV--EGLTVVGQGFGEVQEPGKTFVH--AEFDGILGLTPSLAA- 190  
QY 146 DDLSEFPFDSLVKQTHVNPFLSLQCGAGPFLNQSEVLASVGGSMIIGGDHSLYTGLW 205  
DB 191 -GGVTFVFDNMAQ-----NLVALPM-----FSVYMSNPGSGSELTFGYPDFHSGSLN 241  
QY 206 YTPIRREYVEIIVRVEINGQDLKMDCKEYNDKSIYDSTGNTNLRPKKFEAAVSIK 265  
DB 242 WVPVTKQAWQIALDGIQVG--DSVMFCSE--GCQAVDTGTSLTGP-----PGKIQQLQ 293  
QY 266 AASSTKFPDGLGQLVCHQAGTTPNIPFVLSILMGEVNTQSFRT-----TILPQQ 320  
DB 294 EALGATYVDEGY-----SVQC-----ANLNMMDVT-----FIINGVPTLNPTA 333  
QY 321 YLRPVEDVATSDQDCYKFAISQSSTG-----TVMGAVIMEGYVYVDFRKR 367











GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:17:12 ; Search time 50.29 Seconds  
(without alignments)  
1400.058 Million cell updates/sec

Title: US-09-724-571-58

Perfect score: 2156

Sequence: 1 ETDEPEEPGRGSGFVEMVD.....GPFVTLDMEDCGYNIPOTDE 407

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2156	100.0	532	Q9ULS1	Q9ULS1 homo sapien
2	2001.5	92.8	476	Q9BYC1	Q9BYC1 homo sapien
3	1898	88.0	457	Q9BYC0	Q9BYC0 homo sapien
4	1753.5	81.3	432	Q9BYB9	Q9BYB9 homo sapien
5	1149	53.3	266	Q9CUU5	Q9CUU5 mus musculus
6	1121	52.0	439	Q9H2V8	Q9H2V8 homo sapien
7	1116.5	51.8	514	Q9JL18	Q9JL18 mus musculus
8	967.5	44.9	396	Q9NZL1	Q9NZL1 homo sapien
9	933	43.3	468	Q9NZL2	Q9NZL2 homo sapien
10	712.5	33.0	213	Q9F0D2	Q9F0D2 homo sapien
11	566.5	26.3	255	Q9R1P7	Q9R1P7 mus musculus
12	332.5	15.4	391	Q9VKP6	Q9VKP6 drosophila
13	332	15.4	354	Q9GYX7	Q9GYX7 boophilus m
14	312.5	14.5	386	Q9BGU5	Q9BGU5 bos taurus
15	308	14.3	384	Q9DEC2	Q9DEC2 xenopus lae
16	308	14.3	385	Q9DEC4	Q9DEC4 rana catesb

17	305	14.1	386	6	Q9GMV7
18	305	14.1	387	6	Q9GMV8
19	304.5	14.1	372	5	Q9VLK3
20	304.5	14.1	387	13	Q9DDV5
21	304	14.1	387	6	Q9GMV9
22	303.5	14.1	383	13	Q9DE45
23	303	14.1	383	13	Q9DEC3
24	302.5	14.0	376	13	Q9PUR8
25	301	14.0	382	13	Q9PRG9
26	301	14.0	423	5	Q9VKP7
27	298.5	13.8	384	13	Q91322
28	295.5	13.7	386	6	Q9GMV6
29	294	13.6	396	13	Q93428
30	290.5	13.5	381	6	Q9GK11
31	288	13.4	399	13	Q93458
32	287.5	13.3	444	5	Q21966
33	284	13.2	398	13	P87370
34	284	13.2	427	5	P91802
35	281	13.0	378	13	Q9F0R9
36	279.5	13.0	390	6	Q9GK10
37	278	12.9	370	6	Q9TTW1
38	278	12.9	399	13	Q9DD89
39	278	12.9	422	5	Q96906
40	277.5	12.9	380	6	Q28950
41	277	12.8	446	5	Q9N9H3
42	276	12.8	389	13	Q9FWK1
43	276	12.8	389	13	Q9W643
44	276	12.8	396	13	Q9DEX3
45	275	12.8	385	6	Q29080

#### ALIGNMENTS

RESULT 1

Q9ULS1 PRELIMINARY; PRT; 532 AA.  
 AC Q9ULS1;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE KIAA1149 PROTEIN (FRAGMENT).  
 GN KIAA1149.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=20039618; PubMed=10574461;  
 RA Hirose M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;  
 RT "Characterization of cDNA clones selected by the Genemark analysis  
 RT from size-fractionated cDNA libraries from human brain.";  
 RL DNA Res. 6:329-336(1999).  
 DR EMBL; AB032975; BAA86463.2; .  
 DR HSSP; P56272; 1AM5.  
 DR MEROPS; A01.004; .  
 DR InterPro; IPR001969; Asp\_protease.  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp; 1.  
 DR PRINTS; PR00792; PEPsin.  
 DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
 FT NON\_TER 1  
 SQ SEQUENCE 532 AA; 58720 MW; 98B135D0D5FBD2E8 CRC64;

Query Match 100.0%; Score 2156; DB 4; Length 532;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-175;  
 Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSGFVEMVDNLKSGGYVEMTVGSPQTLNLLVDTGGSNFAVGAAP 60  
 |||||

Db 77 ETDEEPEEPCRRGSFVEMVDNLKSGSQGYVYEMTVGSPQTLNLLVDTGSSNFAVGAAP 136  
QY 61 HPFLHRYQRLSSTYRDLRKGVYVPTQGWEGELGTDLSIPHGPNTVVRANIAAITE 120  
Db 137 HPFLHRYQRLSSTYRDLRKGVYVPTQGWEGELGTDLSIPHGPNTVVRANIAAITE 196  
QY 121 SDRFFINGSNWEGILGLAYAEIARPDSDLSLVKQTHVFNLSQLCAGFPLNQS 180  
Db 197 SDRFFINGSNWEGILGLAYAEIARPDSDLSLVKQTHVFNLSQLCAGFPLNQS 256  
QY 181 EVLASVGSMIIGDHSLYTGSWTPPIRREWYEVIIIVRVEINGODLKMDCKEYNYDK 240  
Db 257 EVLASVGSMIIGDHSLYTGSWTPPIRREWYEVIIIVRVEINGODLKMDCKEYNYDK 316  
QY 241 SIYDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFVWLGEOVLCVQAGTTPWNIFFVIS 300  
Db 317 SIYDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFVWLGEOVLCVQAGTTPWNIFFVIS 376  
QY 301 LYLMEVNTQSFRITILPQOYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYV 360  
Db 377 LYLMEVNTQSFRITILPQOYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYV 436  
QY 361 FDRARKRIGFAVSACHVHDEFTAAVEGPFVTLDMEDCGYNIPQTD 407  
Db 437 FDRARKRIGFAVSACHVHDEFTAAVEGPFVTLDMEDCGYNIPQTD 483  
RESULT 2  
Q9BYC1 PRELIMINARY; PRT; 476 AA.  
ID Q9BYC1  
AC Q9BYC1: 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE BETA-SITE APP CLEAVING ENZYME I-476.  
GN BACE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRIN;  
RX MEDLINE=21408467; PubMed=11516562;  
RA Tanahashi H., Tabira T.;  
RT "Three novel alternatively spliced isoforms of the human beta-site APP  
cleaving enzyme (BACE) and their effect on amyloid beta-peptide  
production.";  
RL Neurosci. Lett. 307:9-12(2001).  
DR EMBL; AB050436; BAB40931.1; -.  
DR HSSP; P32329; LYPS.  
DR InterPro; IPR001969; Asp\_protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF00026; asp; 1.  
DR PRINTS; PR00792; PEPSIN.  
DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN.1.  
SQ SEQUENCE 476 AA; 52907 MW; 6C8C87F8A953AF66 CRC64;

Query Match 92.8%; Score 2001.5; DB 4; Length 476;  
Best Local Similarity 93.9%; Pred. No. 4.le-162;  
Matches 382; Conservative 0; Mismatches 0; Indels 25; Gaps 1;  
QY 1 ETDEEPEEPCRRGSFVEMVDNLKSGSQGYVYEMTVGSPQTLNLLVDTGSSNFAVGAAP 60  
Db 46 ETDEEPEEPCRRGSFVEMVDNLKSGSQGYVYEMTVGSPQTLNLLVDTGSSNFAVGAAP 105  
QY 61 HPFLHRYQRLSSTYRDLRKGVYVPTQGWEGELGTDLSIPHGPNTVVRANIAAITE 120  
Db 106 HPFLHRYQRLSSTYRDLRKGVYVPTQGWEGELGTDLSIPHGPNTVVRANIAAITE 165  
QY 121 SDRFFINGSNWEGILGLAYAEIARPDSDLSLVKQTHVFNLSQLCAGFPLNQS 180  
|||||

Db 166 SDKFFINGSNWEGILGLAYAEIARPDSDLSLVKQTHVFNLSQLCAGFPLNQS 200  
QY 181 EVLASVGSMIIGDHSLYTGSWTPPIRREWYEVIIIVRVEINGODLKMDCKEYNYDK 240  
Db 201 EVLASVGSMIIGDHSLYTGSWTPPIRREWYEVIIIVRVEINGODLKMDCKEYNYDK 260  
QY 241 SIYDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFVWLGEOVLCVQAGTTPWNIFFVIS 300  
Db 261 SIYDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFVWLGEOVLCVQAGTTPWNIFFVIS 320  
QY 301 LYLMEVNTQSFRITILPQOYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYV 360  
Db 321 LYLMEVNTQSFRITILPQOYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYV 380  
QY 361 FDRARKRIGFAVSACHVHDEFTAAVEGPFVTLDMEDCGYNIPQTD 407  
Db 381 FDRARKRIGFAVSACHVHDEFTAAVEGPFVTLDMEDCGYNIPQTD 427  
RESULT 3  
Q9BYC0 PRELIMINARY; PRT; 457 AA.  
ID Q9BYC0  
AC Q9BYC0: 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE BETA-SITE APP CLEAVING ENZYME I-457 (BETA-SITE APP CLEAVING ENZYME  
TYPE C).  
GN BACE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRIN;  
RX MEDLINE=21408467; PubMed=11516562;  
RA Tanahashi H., Tabira T.;  
RT "Three novel alternatively spliced isoforms of the human beta-site APP  
cleaving enzyme (BACE) and their effect on amyloid beta-peptide  
production.";  
RL Neurosci. Lett. 307:9-12(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EXOCRINE PANCREAS;  
RA Zaccchetti D., De Pietri Tonelli D., Schurbus R.;  
RT "New beta-site APP cleaving enzyme isoform (BACE-1C) obtained from  
human pancreas.";  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB050437; BAB40932.1; -.  
DR HSSP; P32329; LYPS.  
DR InterPro; IPR001969; Asp\_protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF00026; asp; 1.  
DR PRINTS; PR00792; PEPSIN.  
DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN.1.  
SQ SEQUENCE 457 AA; 51068 MW; C794C9A9E85FE7A2 CRC64;

Query Match 88.0%; Score 1898; DB 4; Length 457;  
Best Local Similarity 89.2%; Pred. No. 2.5e-153;  
Matches 363; Conservative 0; Mismatches 0; Indels 44; Gaps 1;  
QY 1 ETDEEPEEPCRRGSFVEMVDNLKSGSQGYVYEMTVGSPQTLNLLVDTGSSNFAVGAAP 60  
Db 46 ETDEEPEEPCRRGSFVEMVDNLKSGSQGYVYEMTVGSPQTLNLLVDTGSSNFAVGAAP 105  
QY 61 HPFLHRYQRLSSTYRDLRKGVYVPTQGWEGELGTDLSIPHGPNTVVRANIAAITE 120  
Db 106 HPFLHRYQRLSSTYRDLRKGVYVPTQGWEGELGTDLSIPHGPNTVVRANIAAITE 145  
QY 121 SDRFFINGSNWEGILGLAYAEIARPDSDLSLVKQTHVFNLSQLCAGFPLNQS 180  
|||||

```

Db 146 -----PDSLEFFSLVKQTHVPLNSLQCGAGFLNQ 181
QY 181 EVLASVGSMIGIDHSYTGSLWYTPIRREWYEVIIVRVEINGQDLKMDCKEYNDK 240
Db 182 EVLASVGSMIGIDHSYTGSLWYTPIRREWYEVIIVRVEINGQDLKMDCKEYNDK 241
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLGQVLCWQAGTTPWNIPFV 300
Db 242 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLGQVLCWQAGTTPWNIPFV 301
QY 301 LYLMGEVNTQSFRTITLPOQYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYV 360
Db 302 LYLMGEVNTQSFRTITLPOQYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYV 361
QY 361 FDRARRKRGFAVSAHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDTE 407
Db 362 FDRARRKRGFAVSAHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDTE 408

RESULT 4
Q9BYB9
ID Q9BYB9 PRELIMINARY; PRT; 432 AA.
AC Q9BYB9;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE BETA-SITE APP CLEAVING ENZYME I-432.
GN BACE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=21408467; PubMed=11516562;
RA Tanahashi H., Tabira T.;
RT "Three novel alternatively spliced isoforms of the human beta-site APP
RT cleaving enzyme (BACE) and their effect on amyloid beta-peptide
RT production.";
RL Neurosci. Lett. 307:9-12(2001).
DR EMBL; AB050438; BAB40933.1; -.
DR HSSP; P32329; 1YPS.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN.1.
SQ SEQUENCE 432 AA; 48212 MW; 96FC81EGF0EED01B CRC64;

Query Match 81.3%; Score 1753.5; DB 4; Length 432;
Best Local Similarity 83.0%; Pred. No. 4.8e-141;
Matches 338; Conservative 0; Mismatches 0; Indels 69; Gaps 1;

QY 1 ETDEPEPPGRGFSFVEMVNDLRKSGQYVEMTVGSPPTNLINLVDTGSSNFVGAAP 60
Db 46 ETDEPEPPGRGFSFVEMVNDLRKSGQYVEMTVGSPPTNLINLVDTGSSNFVGAAP 105
QY 61 HPFLHRYQRLSTYRDLRGVYVPTQKWEGLGTDLDVSIHPGNVTVRAIAAITE 120
Db 106 HPFLHRYQRLSTYRDLRGVYVPTQKWEGLGTDL----- 145
QY 121 SDKFFINGSNWEGILGAYAEIARPDSDLEPFDSLKVQTHVPLNSLQCGAGFLPNOS 180
Db 146 -----LCGAGFLPNOS 156
QY 181 EVLASVGSMIGIDHSYTGSLWYTPIRREWYEVIIVRVEINGQDLKMDCKEYNDK 240
Db 157 EVLASVGSMIGIDHSYTGSLWYTPIRREWYEVIIVRVEINGQDLKMDCKEYNDK 216
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLGQVLCWQAGTTPWNIPFV 300

```

```

Db 217 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLGQVLCWQAGTTPWNIPFV 276
QY 301 LYLMGEVNTQSFRTITLPOQYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYV 360
Db 277 LYLMGEVNTQSFRTITLPOQYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYV 336
QY 361 FDRARRKRGFAVSAHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDTE 407
Db 337 FDRARRKRGFAVSAHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDTE 383

RESULT 5
Q9CUU5
ID Q9CUU5 PRELIMINARY; PRT; 266 AA.
AC Q9CUU5;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE BETA-SITE APP CLEAVING ENZYME (FRAGMENT).
GN BACE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=BRAIN;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Cariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK014390; BAB29317.1; -.
DR MEROPS; A01.004; -.
DR MGD; MGI:1346542; Bace.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 2.
FT NON_TER 1
SQ SEQUENCE 266 AA; 30201 MW; B913FDA8ADAB4238 CRC64;

Query Match 53.3%; Score 1149; DB 11; Length 266;
Best Local Similarity 99.1%; Pred. No. 7.7e-90;
Matches 215; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 191 IIGGDHSYTGSLWYTPIRREWYEVIIVRVEINGQDLKMDCKEYNDKSIDVSGTTNL 250
Db 1 IIGGDHSYTGSLWYTPIRREWYEVIIVRVEINGQDLKMDCKEYNDKSIDVSGTTNL 60
QY 251 RLPKVFEEAAVKSIAASSTKFPDGFGLGQVLCWQAGTTPWNIPFVSIYLMGEVTQ 310
Db 61 RLPKVFEEAAVKSIAASSTKFPDGFGLGQVLCWQAGTTPWNIPFVSIYLMGEVTQ 120
QY 311 SFRITILPOQYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYVDFDRARRKRG 370
Db 121 SFRITILPOQYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYVDFDRARRKRG 180

```

DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)	101
DE	ASPARTYL PROTEASE 1.	102
GN	BACE2.	103
OS	Mus musculus (Mouse).	104
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	105
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	106
OX	NCBI_TaxID=10090;	107
RN	[1]	108
RP	SEQUENCE FROM N.A.	109
RA	Choi D.K., Sugano S., Sakaki Y.;	110
RT	"Molecular characterization of the mouse Aspl gene, a homolog of the	111
RT	human ASPl (Down Syndrome Region aspartyl protease).";	112
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.	113
DR	EMBL: AF216310; AAF36599.1; -.	114
DR	HSSP: P32329; 1YPS.	115
DR	MEROPS: A01_041; -.	116
DR	MGI: 1860440; Bace2.	117
DR	InterPro: IPR001969; Asp_protease.	118
DR	InterPro: IPR001461; Pepsin.	119
DR	Pfam: PF00026; asp; 3.	120
DR	PRINTS: PR00792; PEPSIN.	121
DR	PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_2.	122
KW	Protease. 514 AA; 55799 MW; A70725F2C1DF5B47 CRC64;	123
SQ	SEQUENCE 514 AA; 55799 MW; A70725F2C1DF5B47 CRC64;	124
Query Match 51.8%; Score 1116.5; DB 11; Length 514;		
Best Local Similarity 53.7%; Pred. No. 1.2e-86;		
Matches 209; Conservative 66; Mismatches 109; Indels 5; Gaps 3;		
QY	8 EPCR-RGSFVEMVDNLGKSGQGYVEMTVGSPQTLNLTVDGTSSNFVAGAPHPFLHR 66	125
Db	65 EPVRATANFLAMVDNLQDGSRGYILEMLIGTPQKQVQLVDTGSSNFVAGAPHSYIDT 124	126
QY	67 YYORQLSSTYRDLRKGVYVPTQGWEGELGDLVSIPIHGPVNTVRANAAITESDKFFI 126	127
Db	125 YFDSSESTYHSGDFVTYKVTGSGMTGFGVDELVTIPKGFNSFLVNATIFESENFFL 184	128
QY	127 NGSMWEGILGATVATVAPDDSLPPFDSLVKQTHVPLNLSLQCGAGFLNQSEVLAV 186	129
Db	185 PGIKWNGILGLAYAAALAKPSSSLETFEDSLVAQAKIPDIFSMQMGAGLPVAGS---GTN 241	130
QY	187 GGSMTIGGIDHSLTGSLWYTPIRREWYEVIIIVRVEINGODLKMCKEYNYDKSIVDSG 246	131
Db	242 GGSVLGGLIEPSLYKGDWYTPKEEYQYIEILKLEIGGNLMDCREYNADKAIIVDSG 301	132
QY	247 TTNLRPKKVEAAVKSIAASSTEKFPDGLGEQLVCWQAGTTPWNIPFVISLYLMGE 306	133
Db	302 TLLRLPKVFDVAVVEAVARTSLIPEFSDGFWTGAQACWTNSETPNAYFPKISILRDE 361	134
QY	307 VTNOSFRITILPQOYLRPVEDVATSDQCYKFAISQSSTGTVMGAVIMEGFYVVFDRAR 366	135
Db	362 NASRFRITILPOLYIQPMMGAGFNY-ECYRFGISSSTNALVIGATVMGEGFYVVFDRAR 420	136
QY	367 RIGFAVACHVHDEPRTAAVEGPPVFLDM 395	137
Db	421 RVGFAVSPCAETEGTIVSEISGPFSTEDI 449	138
RESULT 8		
Q9NZL1	PRELIMINARY; PRT; 396 AA.	
AC Q9NZL1;		
DT 01-OCT-2000 (TReMBLrel. 15, Created)		
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)		
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)		
DE ASPARTYL PROTEASE.		
GN BACE2.		
OS Homo sapiens (Human).		
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX NCBI_TaxID=9606;		
RN [1]		

DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)	407
DE	ASPARTYL PROTEASE 1.	408
GN	BACE2.	409
OS	Mus musculus (Mouse).	410
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	411
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	412
OX	NCBI_TaxID=10090;	413
RN	[1]	414
RP	SEQUENCE FROM N.A.	415
RA	Choi D.K., Sugano S., Sakaki Y.;	416
RT	"Molecular characterization of the mouse Aspl gene, a homolog of the	417
RT	human ASPl (Down Syndrome Region aspartyl protease).";	418
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.	419
DR	EMBL: AF216310; AAF36599.1; -.	420
DR	HSSP: P32329; 1YPS.	421
DR	MEROPS: A01_041; -.	422
DR	MGI: 1860440; Bace2.	423
DR	InterPro: IPR001969; Asp_protease.	424
DR	InterPro: IPR001461; Pepsin.	425
DR	Pfam: PF00026; asp; 3.	426
DR	PRINTS: PR00792; PEPSIN.	427
DR	PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_1.	428
KW	Protease. 514 AA; 55799 MW; A70725F2C1DF5B47 CRC64;	429
SQ	SEQUENCE 514 AA; 55799 MW; A70725F2C1DF5B47 CRC64;	430
Query Match 52.0%; Score 1121; DB 4; Length 439;		
Best Local Similarity 54.8%; Pred. No. 3.9e-87;		
Matches 207; Conservative 63; Mismatches 104; Indels 4; Gaps 2;		
QY	18 MVDNLGKSGQGYVEMTVGSPQTLNLTVDGTSSNFVAGAPHPFLHRYORQLSSYR 77	431
Db	1 MVDNLQDGSRGYILEMLIGTPQKQVQLVDTGSSNFVAGAPHSYIDTIFDTERSSYR 60	432
QY	78 DLKRGVYVPTQGWEGELGDLVSIPIHGPVNTVRANAAITESDKFFINGWGLGL 137	433
Db	61 SKGFDVTYKVTGSGMTGFGVDELVTIPKGFNTSFLVNATIFESENFFLPKIKWNGILGL 120	434
QY	138 AYAEIARPDPSLEPFDSLVKQTHVPLNLSLQCGAGFLNQSEVLAVSGSMIIGDIDH 197	435
Db	121 AYATLAKPSSSLETFEDSLVQANIPNVFSMQMGAGLPVAGS---GTMGSLVLGGIEP 177	436
QY	198 SLYTGSLSWYTPIRREWYEVIIIVRVEINGODLKMCKEYNYDKSIVDSGTTNLRPKKVF 257	437
Db	178 SLYKGDWYTPKEEYQYIEILKLEIGGSLNLDREYNADKAIIVDSGTTNLRPKKVF 237	438
QY	258 EAAVKSIAASSTEKFPDGLGEQLVCWQAGTTPWNIPFVISLYLMGEVTNOSFRITIL 317	439
Db	238 DAVVEAVARASLIPEFSDGFWTGAQACWTNSETPNAYFPKISILRDESNSSFRITIL 297	440
QY	318 PQOYLRPVEDVATSDQCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVACHV 377	441
Db	298 POLYIQPMMGAGFNY-ECYRFGISSSTNALVIGATVMGEGFYVVFDRARKRIGFAVACHV 356	442
QY	378 HDEFTAAVEGPPVFLDM 395	443
Db	357 IAGRAVSEISGPFSTEDV 374	444
RESULT 7		
Q9JL18	PRELIMINARY; PRT; 514 AA.	
AC Q9JL18;		
DT 01-OCT-2000 (TReMBLrel. 15, Created)		
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)		

```
RP SEQUENCE FROM N.A.
RX MEDLINE-20422477; PubMed-10965118;
RA Solans A., Estivill X., de La Luna S.;
RT "A new aspartyl protease on 21q22.3; BACE2, is highly similar to
RL Alzheimer's amyloid precursor protein beta-secretase.";
RL Cytochrome. Cell Genet. 89:177-184(2000).
DR EMBL; AF188277; AAF35836.1; -.
DR HSSP; P00797; 2REN.
DR MEROPS; A01.041; -.
DR InterPro; IPR001969; Asp_protease.
DR Pfam; PF00026; asp; 2.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_2.
KW Protease.
SQ SEQUENCE 396 AA; 43013 MW; 5023A7AF391CEAC9 CRC64;

Query Match 44.9%; Score 967.5; DB 4; Length 396;
Best Local Similarity 54.6%; Pred. No. 4e-74;
Matches 183; Conservative 54; Mismatches 85; Indels 13; Gaps 4;

QY 9 PGRGSEFVEMVNDLRGKSGQGYVEMVVGSPPTNLILVDTGSSNFVAGAAPHPFLHRY 68
Db 71 PAGAANFLAMVDNLQDGRGYYLEMLIGTPQKQLIILVDTGSSNFVAGTPHSDYIDTF 130
QY 69 QRLSSTYRDLRKGVVYPYTOGKWEGLGTLVSIPIHGPNTVTRANIAATESDKFFNG 128
Db 131 DTERSSTYRSKGFDTVKYTOGSMGTGFGEDLVTPKGFNTSFLVNIATIFESNFPLG 190
QY 129 SNWEGILGLAVAEIARPDLSLEPFDSLVKQTHVNPVLFSLQCGAGPLNQSEVLASVG 188
Db 191 IKWNGILGLATATLAKPSSLETFFDSLVQTQANIPNVFMSQMCGAGLPVAGS---GTNGG 247
QY 189 SMIIIGDHSLSYTGSLWYTPIRREWYEVILVIRVEINGODLKMCKEYNDKSIDVSGTT 248
Db 248 SLVGGIEPSLYKGDWYTPKEWYQIETLKEIGGQSLNDCREYNADKAIVDSGTT 307
QY 249 NLRPKVFEAAVKSIRKASTETKFPDGLVQWQAGTTPWNIFFPVSILYMGVET 308
Db 308 LLRLPQKVFDAVAVARASLPEFSDGFWTGSOLACNTSETPSWYFPKISIVLRDENS 367
QY 309 NOSFRITLPQYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYVYFDRARKRI 368
Db 368 SRSFRITLPQK-LRVLQ-----CLKPFGLSQ 393

RESULT 9
Q9NZL2 PRELIMINARY; PRT; 468 AA.
AC Q9NZL2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ASPARTYL PROTEASE.
GN BACE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CORD BLOOD;
RA Zhang Q.H., Ye M., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
RA Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
RT "Human partial CDS cloned from cd34+ stem cells.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF161367; AAF28927.1; -.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 1.
FT NON_TER 1
SQ SEQUENCE 213 AA; 24338 MW; EC9D3FA31CFA835C CRC64;

Query Match 33.0%; Score 712.5; DB 4; Length 213;
Best Local Similarity 83.5%; Pred. No. 8.7e-53;
Matches 137; Conservative 4; Mismatches 12; Indels 11; Gaps 1;

QY 193 GGIDHSLSYTGSLWYTPIRREWYEVILVIRVEINGODLKMCKEYNDKSIDVSGTTNRL 252
Db 1 GGIDHSLSYTGSLWYTPIRREWYEVILVIRVEINGODLKMCKEYNDKSIDVSGTTNRL 60
```

```
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_2.
KW Protease.
SQ SEQUENCE 468 AA; 50324 MW; 717E0920126A0142 CRC64;

Query Match 43.3%; Score 933; DB 4; Length 468;
Best Local Similarity 46.5%; Pred. No. 4.5e-71;
Matches 180; Conservative 59; Mismatches 94; Indels 54; Gaps 3;

QY 9 PGRGSEFVEMVNDLRGKSGQGYVEMVVGSPPTNLILVDTGSSNFVAGAAPHPFLHRY 68
Db 71 PAGAANFLAMVDNLQDGRGYYLEMLIGTPQKQLIILVDTGSSNFVAGTPHSDYIDTF 130
QY 69 QRLSSTYRDLRKGVVYPYTOGKWEGLGTLVSIPIHGPNTVTRANIAATESDKFFNG 128
Db 131 DTERSSTYRSKGFDTVKYTOGSMGTGFGEDLVTPKGFNTSFLVNIATIFESNFPLG 190
QY 129 SNWEGILGLAVAEIARPDLSLEPFDSLVKQTHVNPVLFSLQCGAGPLNQSEVLASVG 188
Db 191 IKWNGILGLATATLAKPSSLETFFDSLVQTQANIPNVFMSQMCGAGLPVAGS---GTNGG 247
QY 189 SMIIIGDHSLSYTGSLWYTPIRREWYEVILVIRVEINGODLKMCKEYNDKSIDVSGTT 248
Db 248 SLVGGIEPSLYKGDWYTPKEWYQIETLKEIGGQSLNDCREYNADKAIVDSGTT 307
QY 249 NLRPKVFEAAVKSIRKASTETKFPDGLVQWQAGTTPWNIFFPVSILYMGVET 308
Db 308 LLRLPQKVFDAVAVARASLPEFSDGFWLQWQAGTTPWNIFFPVSILYMGVET 376
QY 369 GFAVSACHVHDEPRTAAVEGPFVTLDM 395
Db 376 GFAVSACHVHDEPRTAAVEGPFVTLDM 403

RESULT 10
Q9P0D2 PRELIMINARY; PRT; 213 AA.
AC Q9P0D2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HSPCL04 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CORD BLOOD;
RA Zhang Q.H., Ye M., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
RA Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
RT "Human partial CDS cloned from cd34+ stem cells.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF161367; AAF28927.1; -.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 1.
FT NON_TER 1
SQ SEQUENCE 213 AA; 24338 MW; EC9D3FA31CFA835C CRC64;

Query Match 33.0%; Score 712.5; DB 4; Length 213;
Best Local Similarity 83.5%; Pred. No. 8.7e-53;
Matches 137; Conservative 4; Mismatches 12; Indels 11; Gaps 1;

QY 193 GGIDHSLSYTGSLWYTPIRREWYEVILVIRVEINGODLKMCKEYNDKSIDVSGTTNRL 252
Db 1 GGIDHSLSYTGSLWYTPIRREWYEVILVIRVEINGODLKMCKEYNDKSIDVSGTTNRL 60
```

```

QY 253 PKKVEAAVKSIIKAASSTKFPDGLGQLVCMQAGTTPWNIFFVISLYLMGEVTVNSF 312
Db 61 PKKVEAAVKSIIKAASSTKFPDGLGQLVCMQAGTTPWNIFFVISLYLMGEVTVNSF 120
QY 313 RIIVLPQYLRPVEDVATSDODCYKFAISQSSTGTGVGAVIMEG 356
Db 121 RIIVLPQYLRP-----WKMPRPKTTVTVCCHLTIVHG 153

RESULT 11
Q9RLP7 PRELIMINARY; PRT; 255 AA.
AC Q9RLP7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ASPARTYL PROTEASE (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RA Accarino M., Fumagalli P., Taramelli R., Ottolenghi S.;
RT "Cloning of a gene from chromosome 21 Down Region encoding a potential
RT transmembrane protease.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF051150; AAD45964.1; -.
DR MEROPS: A01.041; -.
DR InterPro: IPR001969; Asp_protease.
DR Pfam: PF000026; asp; 2.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_1.
KW protease.
FT NON_TER 1
SQ SEQUENCE 255 AA; 28685 MW; 53DE317815996D63 CRC64;

Query Match 26.3%; Score 566.5; DB 11; Length 255;
Best Local Similarity 53.4%; Pred. No. 3.1e-40;
Matches 102; Conservative 35; Mismatches 53; Indels 1; Gaps 1;

QY 205 WYTPTRREWYVIVRVVEINGDLKMDCKEYNDKSIDVSGTTLNLRPKKVFEEAVKSI 264
Db 1 WYTPKEEYVYIETLKLKGQNLNLCREYNADKAIVDSGTTLLRLPQKVFDAVVEAV 60
QY 265 KAASSTKFPDGLGQLVCMQAGTTPWNIFFVISLYLMGEVTVNQSFRTILPQOYLRP 324
Db 61 ARTSLIPEFSDGFWTGAQLACWTNSETPWAYFPKISIVLRDENASRFRITILPOLYIOP 120
QY 325 VEDVATSDODCYKFAISQSSTGTGVGAVIMEGFYVYVDFRARKRIGFAYVSACHVHDFRTA 384
Db 121 MMGAGFNY-ECYRFGISSTNALVIGATVMEGFYVYVDFRQRRVGFAYSPCAEIEGTTVS 179
QY 385 AVEGPEVTLDL 395
Db 180 EISGPESTEDI 190

RESULT 12
Q9VKP6 PRELIMINARY; PRT; 391 AA.
AC Q9VKP6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CG17134 PROTEIN.
GN CG17134.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
```

```

NX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej D.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flossler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Nusskern D.R., Pacleb J.M.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR EMBL: AE003630; AAF53016.1; -.
DR HSP: P00794; 4CMS.
DR MEROPS: A01.0PW; -.
DR FlyBase: FBgn0032304; CG17134.
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR001461; Pepsin.
DR Pfam: PF000026; asp; 1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; 2.
DR Aspartyl protease; Hydrolase.
KW SEQUENCE 391 AA; 42016 MW; 99A18E7131025E1B CRC64;

Query Match 15.4%; Score 332.5; DB 5; Length 391;
Best Local Similarity 29.1%; Pred. No. 5.1e-20;
Matches 105; Conservative 61; Mismatches 146; Indels 49; Gaps 13;

QY 20 DNLRGKGGQYVEMTVGSPPTLNLIVDTGSSNFVAGAAPHPF-----LHRYQRQLS 73
Db 66 ENLHNSMNEYGVIAIGTPEQRNFILFTGSAANLWPSASCPASNTACQRHNYDSAS 125
QY 74 STYDLRGVYVYPTQCKWEGELGTLVSIPIHGNVTVVRANLAATESDKFFINGSNWEG 133
Db 126 STYVANGEEFAEYGTGSLSGFLSNDIVTIA-GISIQNTGEALSEPCTFVD-APFAG 183
QY 134 ILGLAYAEIARPDSDLEPFDFSLVKQTHVPN-LFSLQLCGAGFPLNQSVLASVGGSMII 192
Db 184 ILGLAFSAIA--VDGVTTPFDNMISQGLLDEPVISFYLRKQG-----TAVRGELIL 233
```



```
QY 193 GGIDHSLYGTSLWYTPIRREWYEVILVIRVEINGODLMDCKEYNDKSIYDVGTTNLR 252
||||| ||| ||| ||| ||| : : : : : ||| : : : ||| : : : ||| : : : |||
Db 234 GGIDSLYRGSTYVPVSPAYQKVTIKNTGTLNCGC-----QAIDTGTSLIAY 287
QY 253 PKKVEAAVKSKAASSTKEKPDGFWLGEOLY-CWAGTTPWNIPFPVLSILYMGVETNQS 311
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 288 PLAAVYRKINRQLGATDND-----GEAFVRCGRVSS-----LPKYNLNLGGTV---- 329
QY 312 FRITILPQOYLPRVEDVATSDQDCKY-FAISQSSTGTVWGVIMEGFYVVDRAKRKRGF 370
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 330 --FTLAPRDYI--VKVTONGOTYCMSAFTYMEGLSFWILGDVFIKGYTFVDFKGNERTGF 385
QY 371 A 371
|
Db 386 A 386

RESULT 13
Q9GYX7 PRELIMINARY; PRT; 354 AA.
AC Q9GYX7;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE HEME-BINDING ASPARTIC PROTEINASE (FRAGMENT).
OS Boophilus microplus (Cattle tick).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Parasitiformes; Ixodida; Ixodidae; Boophilus.
OX NCBI_TaxID=6941;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PORTO ALEGRE; TISSUE=OVARY;
RA Sorgine M.H.F., Logullo C., Zingali R.B., Paiva-Silva G.O.,
RA Juliano L., Oliveira P.L.;
RT "A heme-binding aspartic proteinase from the eggs of the hard tick
RT Boophilus microplus."
RL J. Biol. Chem. 0:0-0(2000).
DR EMBL; AF286865; AAG00993.1;
DR HSSP; P00797; 2REN.
DR MEROPS; A01.054; -.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp.1.
DR PRINTS; PR00792; PEPsin.
FT NON_TER 1
FT SEQUENCE 354 AA; 38512 MW; B50D7C638CF27091 CRC64;
```

```
Query Match 15.4%; Score 332; DB 5; Length 354;
Best Local Similarity 26.2%; Pred. No. 4.9e-20;
Matches 101; Conservative 74; Mismatches 135; Indels 76; Gaps 15;

QY 19 VDNLRGK-----SGQYVYVEMTVGSPPTLNLIVDTGSSNFAVGAAPHFL- 64
| : : : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 13 VTEIRGALGDPIILNTNNMQFYGIITGTPQSFKLLMDTGSSNFWV-----PSIN 67
QY 65 -----HRYQRLSSYTRDLRKGVYVYQGWEGELGTLVSIPIHGPN-----VTVR 112
| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 68 CDOSMACRDHAKYDSSKSTFTKSGRIRYSGGVVRGITSIDNVGV--GPATVTQYKF 125
QY 116 AAITESKFFINGSNWEGILGLAYAEIARPDSDLPEPFDLSLVKQTHVPN-LFSQLCCAG 174
| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 126 AEMHSDGKLFRAKYDIGIFGLAFPSLSQ--NNQLPFLDAMVQGVVRQAVFSLYL--SK 181
QY 175 FPLNQSEVLASVGSMTIGGIDHSLYGTSLWYTPIRREWYEVILVIRVEINGODLKM-DC 233
| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 182 QPSEQN-----GGEIYFGGINAQRVTGAIHYVPVSOAAHWQVMDNINVGQTTLCVGGC 235
QY 234 KEYNDKSIYDVGTTNLRIPKPKVFEAAVKSIAASSTKEKPDGFWLGEOLVQWAGTTPW 293
| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 236 -----PTVVDSDGTSLSLGP-----SADVETLNRVIGATKTAAGY-----FEVNCATI 277
QY 294 NIFPVISLYLMGE---VTNQSFRTILPQOYLPRVEDVATSDQDCKYFAISQSSTGT--- 347
```

```
Db 278 SSLPITNLNGKSPLOQEAYTIRI-----PLTTGEOCFTRISEDSASGTNLW 327
QY 348 WGVAVIMEGFYVVDRAKRKRGFAVS 373
Db 328 ILGAVFTQTYTVFDRAQNRVGFATA 353

RESULT 14
Q9BGU5 PRELIMINARY; PRT; 386 AA.
AC Q9BGU5;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE CATHEPSIN D (FRAGMENT).
DE CATHEPSIN D (FRAGMENT).
GN CAT-D.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Higuchi M., Miyashita N., Nagamine Y., Awata T.;
RT "Complementary DNA sequence of bovine cathepsin D."
CC Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RL -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR EMBL; AB05312; BAB21620.1; -.
DR HSSP; P07339; ILVB.
DR MEROPS; A01.009; -.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp.1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Aspartyl protease; Hydrolase.
FT NON_TER 1
FT SEQUENCE 386 AA; 41965 MW; 33BF00B80585490D CRC64;

Query Match 14.5%; Score 312.5; DB 6; Length 386;
Best Local Similarity 28.0%; Pred. No. 2.5e-18;
Matches 113; Conservative 69; Mismatches 131; Indels 91; Gaps 20;

QY 8 EPG-RRGSFVEMVDNLGRKSGQYVYVEMTVGSPPTLNLIVDTGSSNFAVGAAPHFL-- 64
Db 35 EPAVRQGPPIPELLKNYMDAQ---YGEIGIGITPPQCFTVVFDGTSANLWVPSIHKLLDI 91
QY 65 ---HRYQRLSSYTRDLRKGVY--VPYQGWEGELGTLVSIPIHGPN-----VTVR 112
| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 92 ACWTHRYNSDKSTY--VKNGTTFDIHGSGSLGSLGSLQSDTVSVPCNPSSSPGGVTVQ 149
QY 113 ANI--AAITESKFFINGSNWEGILGLAYAEIARPDSDLPEPFDLSLVKQTHV-PNLFSLQ 169
| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 150 RQTFGEAIKPGVVFI-AAKFGGILGMAYPRIS--VNNVLPVFDNLMOQLVKNVFS-- 204
QY 170 LCAGAPLNQSEVLASVGSMTIGGIDHSLYGTSLWYTPIRREWYEVILVIRVEINGQDL 229
| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 205 -----FFLNR-DPKAOPGGELMLGTDTSKYRGLSFHNVTRQAYWQIHMQLDQV-GSSL 257
QY 230 KMDCKEYNDKSIYDVGTTNLRIPKPKVFEAAVKSIAASSTKEKPDGFWLGEOLVQWAG 289
| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 258 TV-CK--GGEALVDGTSLVGPVEEVELQKAI-----G 290
QY 290 TTPNIFPVISLYLMGEVNTNQSFRTILPQOYLPRVEDVATSDQDCKYFAISQSSTGT- 347
| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 291 AVP-----LIQGEYMIPECKVSSLPQVTVKLGKDYAKSPED-YALKYSQAGTTVC 340
QY 348 -----VMGVIMEGFYVVDRAKRKRGFAVS 374
Db 341 LSGFMGMIDPPPGGPLWILGDVFIGRYTVFDRDQNRVGLAEAA 384
```

us-09-724-571-58.rspt

Search completed: August 7, 2002, 09:17:13  
Job time: 217 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:14:48 ; Search time 22.58 Seconds  
(without alignments)  
440.266 Million cell updates/sec

Title: US-09-724-571-58  
Perfect score: 2156  
Sequence: 1 ETDEPEPEPGRGSEFVEMVD.....GPFVTLDMEDCGYNIPQTDE 407

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents.AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2150	99.7	501	4	US-09-009-191-2
2	2073	96.2	774	4	US-09-009-191-4
3	1132	52.5	518	3	US-08-999-723-2
4	1132	52.5	518	4	US-09-434-427-2
5	1116.5	51.8	514	4	US-09-717-432-2
6	1116.5	51.8	514	4	US-09-912-484-2
7	299.5	13.9	396	1	US-08-208-007A-13
8	299.5	13.9	396	4	US-09-032-523-9
9	295.5	13.7	412	1	US-08-208-007A-12
10	295.5	13.7	412	4	US-08-974-691-4
11	279.5	13.0	458	6	5217891-15
12	273.5	12.7	409	1	US-09-640-305-6
13	273.5	12.7	409	1	US-08-360-673-6
14	273.5	12.7	427	2	US-08-846-021A-8
15	270	12.5	410	1	US-08-088-633-2
16	270	12.5	410	1	US-08-245-756-2
17	270	12.5	410	1	US-08-441-750-2
18	270	12.5	410	2	US-08-441-751-2
19	270	12.5	410	5	PCT-US92-02521-2
20	250	11.6	349	4	US-09-032-523-3
21	227	10.5	398	1	US-08-328-314-2
22	227	10.5	398	1	US-08-731-045-2
23	212	9.8	419	4	US-08-974-691-3
24	210	9.7	397	3	US-09-079-415-2
25	196.5	9.1	430	1	US-08-535-237-2
26	195.5	9.1	427	1	US-07-958-222A-2
27	193	9.0	330	3	US-08-115-753-1

28	193	9.0	419	3	US-08-115-753-2
29	193	9.0	419	3	US-08-115-753-3
30	187	8.7	420	4	US-09-008-271A-4
31	187	8.7	420	4	US-08-974-691-8
32	184.5	8.6	395	1	US-08-723-938-3
33	184.5	8.6	395	2	US-09-080-538-3
34	184	8.5	445	4	US-08-974-691-6
35	184	8.5	451	4	US-08-974-691-2
36	149	6.9	437	4	US-09-353-332-2
37	129.5	6.0	140	3	US-09-211-631-13
38	129.5	6.0	140	4	US-09-265-628-13
39	129.5	6.0	140	4	US-09-001-141-11
40	129.5	6.0	140	4	US-09-532-803-6
41	129.5	6.0	140	4	US-09-653-403-14
42	97	4.5	1030	4	US-09-091-117-2
43	95.5	4.4	280	4	US-09-160-246-14
44	89.5	4.2	1097	2	US-08-680-326-39
45	88	4.1	746	2	US-08-838-219B-6

## ALIGNMENTS

RESULT 1  
US-09-009-191-2  
; Sequence 2, Application US/09009191  
; Patent No. 6319689  
; GENERAL INFORMATION:  
; APPLICANT: POWELL, DAVID  
; APPLICANT: CHAPMAN, CONRAD  
; APPLICANT: MURPHY, KAY  
; APPLICANT: SMITH, TRUDI  
; TITLE OF INVENTION: ASP2  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: RATNER & PRESTIA  
; STREET: P.O. BOX 980  
; CITY: VALLEY FORGE  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19482  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/009,191  
; FILING DATE: 20-JAN-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: UK 9701684.4  
; FILING DATE: 28-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PRESTIA, PAUL F  
; REGISTRATION NUMBER: 23 031  
; REFERENCE/DOCKET NUMBER: GH-70368  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-407-0700  
; TELEFAX: 610-407-0701  
; TELEX: 846169  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 501 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-009-191-2

Query Match 99.7%; Score 2150; DB 4; Length 501;  
Best Local Similarity 99.8%; Pred. No. 9.4e-223;



QY 129 SNWEGILGLAYAEIARDDSLPEFFDSLVKQTHVFNLSLQCGAGFPLNQSEVLASVG 188  
Db 191 IKWNGILGLAYATLAKFSSSTLEFFDSLYTQANIPNVFSQMCGAGLPVAGS---GTNGG 247  
QY 189 SMITGGIDHSLYTGSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIDVSGTT 248  
Db 248 SLVGGIEPSLYKGDWYTPIKEWYQIIEILKEIGGQSLNLDREYNADKAIVDSGTT 307  
QY 249 NLRLPKKVFEEAAVKSIKAASSTKFPDGLWGLQVLCWQAGTTPWNIFFPVISLYLMEV 308  
Db 308 LLRLPKQVDAVVEAVARSLIPEFSDGFWTGSQACWTNSETPNWYFPKISLYLDENS 367  
QY 309 NOSFRITILPQQLRVEDVATSDQCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRI 368  
Db 368 SRSFRITILPOLYIOPMMGAGLNY-ECYRFGISPSTNALVIGATVMEGFYVVFDRARKRV 426  
QY 369 GFAVSACHVHDEFTAAVEGPFVTLDM 395  
Db 427 GFAASPCAEIAGAAVSEISGPFSTEDV 453

RESULT 4  
US-09-434-427-2

; Sequence 2, Application US/09434427

; Patent No. 6162630

; GENERAL INFORMATION:

; APPLICANT: POWELL, DAVID J.

; APPLICANT: SOUTHAN, CHRISTOPHER

; APPLICANT: CHAPMAN, CONRAD G.

; APPLICANT: EVANS, JOANNE R.

; TITLE OF INVENTION: ASP1

; FILE REFERENCE: GH-70262-D1

; CURRENT APPLICATION NUMBER: US/09/434,427

; CURRENT FILING DATE: 1999-11-04

; EARLIER APPLICATION NUMBER: US 08/999,723

; EARLIER FILING DATE: 1997-10-06

; EARLIER APPLICATION NUMBER: UK 962022.9

; EARLIER FILING DATE: 1996-12-14

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 518

; TYPE: PRT

; ORGANISM: HOMO SAPIENS

US-09-434-427-2

Query Match 52.5%; Score 1132; DB 4; Length 518;

Best Local Similarity 54.0%; Pred. No. 4,1e-113;

Matches 209; Conservative 65; Mismatches 109; Indels 4; Gaps 2;

QY 9 PRRGSEFVEMVDNLRGKSGQGYVYVMTVGSPPQTLNIIIVDTGSSNFAVGAAPHPFLHRY 68

Db 71 PAGRAANFLAMVDNLQDGSRGYYLEMLIGTPPKQLIIVDTGSSNFAVAGTPHSHYIDTF 130

QY 69 QRLSSYTRDLRKGVYVPYTGKWEGLGTLVSIHPGNVTVRANTAAATESDKPFNG 128

Db 131 DTERSSTYRSKGFVDVTKYTGQSWTGFVGEDLVTPKGFNTSFLVNIATIFESENFFLP 190

QY 129 SNWEGILGLAYAEIARDDSLPEFFDSLVKQTHVFNLSLQCGAGFPLNQSEVLASVG 188

Db 191 IKWNGILGLAYATLAKFSSSTLEFFDSLVQANIPNVFSQMCGAGLPVAGS---GTNGG 247

QY 189 SMITGGIDHSLYTGSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIDVSGTT 248

Db 248 SLVGGIEPSLYKGDWYTPIKEWYQIIEILKEIGGQSLNLDREYNADKAIVDSGTT 307

QY 249 NLRLPKKVFEEAAVKSIKAASSTKFPDGLWGLQVLCWQAGTTPWNIFFPVISLYLMEV 308

Db 308 LLRLPKQVDAVVEAVARSLIPEFSDGFWTGSQACWTNSETPNWYFPKISLYLDENS 367

QY 309 NOSFRITILPQQLRVEDVATSDQCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRI 368

Db 427 GFAASPCAEIAGAAVSEISGPFSTEDV 453

Db 368 SRSFRITILPOLYIOPMMGAGLNY-ECYRFGISPSTNALVIGATVMEGFYVVFDRARKRV 426  
QY 369 GFAVSACHVHDEFTAAVEGPFVTLDM 395  
Db 427 GFAASPCAEIAGAAVSEISGPFSTEDV 453

RESULT 5

US-09-717-432-2

; Sequence 2, Application US/09717432

; Patent No. 6291223

; GENERAL INFORMATION:

; APPLICANT: ZHU, YUAN

; APPLICANT: LI, XIAOTONG

; APPLICANT: CHRISTIE, GARY

; APPLICANT: POWELL, DAVID J.

; TITLE OF INVENTION: Mouse Aspartic Secretase-1 (mASPL)

; FILE REFERENCE: GP-70663

; CURRENT APPLICATION NUMBER: US/09/717,432

; CURRENT FILING DATE: 2000-11-21

; PRIOR APPLICATION NUMBER: 60/166,974

; PRIOR FILING DATE: 1999-11-23

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 514

; TYPE: PRT

; ORGANISM: MUS MUSCULUS

US-09-717-432-2

Query Match 51.8%; Score 1116.5; DB 4; Length 514;

Best Local Similarity 53.7%; Pred. No. 1.9e-111;

Matches 209; Conservative 66; Mismatches 109; Indels 5; Gaps 3;

QY 8 EPGR-RGSFVEMVDNLRGKSGQGYVYVMTVGSPPQTLNIIIVDTGSSNFAVGAAPHPFLHR 66

Db 65 EPVRATANFLAMVDNLQDGSRGYYLEMLIGTPPKQLIIVDTGSSNFAVAGAPHSYIDT 124

QY 67 YQORLSSTYTRDLRKGVYVPYTGKWEGLGTLVSIHPGNVTVRANTAAATESDKFFI 126

Db 125 YFDSSESSTYHSKGFVDVTKYTGQSWTGFVGEDLVTPKGFNTSFLVNIATIFESENFFL 184

QY 127 NGSNWEGILGLAYAEIARDDSLPEFFDSLVKQTHVFNLSLQCGAGFPLNQSEVLASV 186

Db 185 PGKWNGLGLAYALAKPSSSLETFEFLVAQAQIPDIFSMQMGAGLPVAGS---GTN 241

QY 187 GGSMTIGGDHSLYTGSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIDVSG 246

Db 242 GGSVLVGGIEPSLYKGDWYTPIKEWYQIIEILKEIGQSLNLDREYNADKAIVDSG 301

QY 247 TTNLRPKKVFEEAAVKSIKAASSTKFPDGLWGLQVLCWQAGTTPWNIFFPVISLYLME 306

Db 302 TTLRLPQVDAVVEAVARTSLIPEFSDGFWTGSQACWTNSETPNWYFPKISLYLME 361

QY 307 VTNQSFRTILPQQLRVEDVATSDQCYKFAISQSSTGTVMGAVIMEGFYVVFDRARK 366

Db 362 NASRFRITILPOLYIOPMMGAGFNY-ECYRFGISSSTNALVIGATVMEGFYVVFDRAR 420

QY 367 RTGFAVSACHVHDEFTAAVEGPFVTLDM 395

Db 421 RVGFAVSPCAEIEGTTVSEISGPFSTEDI 449

RESULT 6

US-09-912-484-2

; Sequence 2, Application US/09912484

; Patent No. 6358725

; GENERAL INFORMATION:

; APPLICANT: Christie, Gary

; APPLICANT: Li, Xiaotong

; APPLICANT: Powell, David J.

; APPLICANT: Zhu, Yuan

```

; TITLE OF INVENTION: Mouse Aspartic Secretase-1 (mASp1)
; FILE REFERENCE: GP-70663-D1
; CURRENT APPLICATION NUMBER: US/09/912,484
; CURRENT FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/166,974
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: 09/717,432
; PRIOR FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 514
; TYPE: PRT
; ORGANISM: MUS MUSCULUS
US-09-912-484-2

Query Match          51.8%; Score 1116.5; DB 4; Length 514;
Best Local Similarity 53.7%; Pred. No. 1.9e-111;
Matches 209; Conservative 66; Mismatches 109; Indels 5; Gaps 3;

QY 8 EPGR-RGSFVEMVNLGRKSGQGYVEMTVGSPPTNLNLVDTGSSNFVAGAPHPFLHR 66
DB 65 EPRATANFLAMVDNLQDSCRGVYLEMLICTPPKQVQILVDTGSSNFVAGAPHSYIDT 124
QY 67 YQROLSTYRDLRKGVYVPYTOGKWEGLCTDLVSIPIHGNVTVVRANIAAITESDKFFI 126
DB 125 YPDESSTYHSKGDVTVKVTQGSWTGFGVEDLVTPKGFNSFLNIATIFSENEFFL 184
QY 127 NSGNNEGILGLAYAIARIPDDSLPEFFDSLVKQTHVPNLFSLQLCGAGFPPLNQSEVLASV 186
DB 185 PGIKNGLILGLAYALAKAPSSLETFFDSLVQAQAKIPDIFSMQMGAGLPVAGS---GTN 241
QY 187 GSGMIIGDHSLYTGSLLWYPIREWTYYEVLIVVEINGODLKMDCKEYNVDKSIYVDSG 246
DB 242 GCSVLGGIEPLSYKGDIIWYTPKEWYQIEILKLEIGGQNLNLDREYNADRAIVDSG 301
QY 247 TTNLRPKKVFVAAKSIKAASSTKFPDGLGQLVCWQAGTTPPWNIFFVISLYLME 306
DB 302 TILLRUPQKVFVAVNAVARTSLIPEFSDGFWTGAQLACWTNSETPWAFYFKISYLRDE 361
QY 307 VTNQSPRITLPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVINMEGFYVDFDARK 366
DB 362 NASRSPRITLPOLYIQPMMGAGFNY-ECYREFGISSTNALVIGATVMEGFYVDFDARQ 420
QY 367 RGFAVSACHVDEFTAAVEGPFVTLDM 395
DB 421 RVGFAVSPCAIEGTTVSEISGPFSTEDI 449

RESULT 7
US-08-208-007A-13
; Sequence 13, Application US/08208007A
; Patent No. 5501969
; GENERAL INFORMATION:
; APPLICANT: HASTINGS, ET AL.
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/208,007A

; TITLE OF INVENTION: Mouse Aspartic Secretase-1 (mASp1)
; FILE REFERENCE: GP-70663-D1
; CURRENT APPLICATION NUMBER: US/09/912,484
; CURRENT FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/166,974
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: 09/717,432
; PRIOR FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 514
; TYPE: PRT
; ORGANISM: MUS MUSCULUS
US-09-912-484-2

Query Match          13.9%; Score 299.5; DB 1; Length 396;
Best Local Similarity 25.9%; Pred. No. 1.2e-23;
Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;

QY 3 DEPEPEPRGRSGFVEMVNLGRKSGQGYVEMTVGSPPTNLNLVDTGSSNFVAGA--- 58
DB 63 DQSAKEP-----LINVLD-----MEYFTGISGPPQNFTVIFDTGSSNLWVPSVYCT 110
QY 59 APHPFLHRYQROLSTYRDLRKGVVPTQGWEGELGTDLSIPIHGNVTVVRANIAAI 118
DB 111 SPACKTHSRFPQSOSSTYSQPGSFSIQYGTGSLGIIGADQVSV-EGLTVVGQOFGESV 169
QY 119 TESDKFFINGSNEGILGLAYAEIARIPDDSLPEFFDSLVKQTHVPNLFSLQLCGAGFPPLN 178
DB 170 TEPQOTFVD-AEPDGIILGLQPSLA--VGGVTPVFONMAQ-----NLVLDLPMF SVYMSSN 222
QY 179 QSEVLASVSGSMIIGDHSLYTGSLLWYTPIRREWYVEVLIVVEINGODLKMDCKEYNY 238
DB 223 PE---GGAGSELGIFGGYDHSFSGSLNWWPVTKQAYWQIADLNQVGG--TVMFCSE--G 275
QY 239 DKSIVDSGTTNLRUPKVFVAAKSIKAASSTKFPDGLGQLVCWQAGTTPWNIFFPV 298
DB 276 QCALVDTGTSLLITGSPDKIKQLQNAIGAAP-----VDGEYAVE-----CANLNVMPD 322
QY 299 ISLYLMGEVNTNQSPRITLPQOYLRPVEDVATSDDCYKFAISQSSTG----- 346
DB 323 VTFTING-----VPYTLSPATY--TLLDFVDGMQFC-----SSGFGGLDIHPGAP 366

QY 347 -TVMGAVIMEGFYVDFDARKRIGFA 371
DB 367 LWILGDVFIROFYSVDFGRGNRVGLA 392

RESULT 8
US-09-032-523-9
; Sequence 9, Application US/09032523
; Patent No. 6232454
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl
; APPLICANT: Baugh, Mariah
; TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
```



Sequence 4, Application US/08974691  
Patent No. 6225103  
GENERAL INFORMATION:  
APPLICANT: Keolsch, Gerald  
APPLICANT: Lin, Xinli  
APPLICANT: Tang, Jordan  
TITLE OF INVENTION: Cloning and Characterization of Napsin  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center, 1201 W. Peachtree  
STREET: St.  
CITY: Atlanta  
STATE: GA  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974.691  
FILING DATE: 20-NOV-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/031,196  
FILING DATE: 20-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/046,126  
FILING DATE: 09-MAY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: OMRF 166  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-873-8794  
TELEFAX: 404-873-8795  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 412 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-974-691-4

Query Match 13.7%; Score 295.5; DB 4; Length 412;  
Best Local Similarity 28.5%; Pred. No. 3.3e-23;  
Matches 105; Conservative 58; Mismatches 146; Indels 59; Gaps 15;  
Qy 30 YVEMTVGSPPTLNILVDTGSSNFAVGAAPHPFL-----HRYQRLSSYRDLKGV 83  
Db 79 YGIGTGPQCTVFTVDTGSSNFWPSIHCKLLDIACVHHKYNDSKSSITVKNKTSF 138  
Qy 84 YVPTQKWEGLTDLVSIPIHGPNTVVRANIAAITESDKFFINGSNWEGI 134  
Db 139 DIHYGSGSLGSLYSQDIPVPCQSASSASALGVKVRQVGEATKPGITFAAKFDGI 198  
Qy 135 LGLAYAEIARPDSDLEPFDSLVKQTHV-PNLFSLQLCGAGFPLNSEVLASVGGSMIIG 193  
Db 199 LGMAVPRIS--VNNVLPVFDNMQKLVDQNFISFYL-----SRDPAQPGGELMLG 248  
Qy 194 GIDHSLYTGSLWYPIREWEYEVILVVEI-NGODLKMDCKEYNDKSIVDSTGNLRL 252  
Db 249 GTDSRYKYGSLSYLNVTKAKWQVHLDOVEVASGLTL--CKE--GCEAIVDTGSLMWG 303  
Qy 253 PKKVFEEAAVKSIKAASSTKFPDGFGLGEQLV-CWQAGTTPWNTFPVISLYLMEVTNQS 311  
Db 304 PVDEYRELQKAIGAVPLIQ-----GEYMIPECVKVST-----LPAITLKLKG-----KG 346

Qy 312 FRITLPQOYLRPVEDVATSDDCYKFAISO-----SSTGTVMGAVIMEGFYVVFDRARK 366  
Db 347 YKLS--PEDYTLKVSQAGKTL-CLSGFMGMDIPPSGPLWILGDVFIQRYTYVFDNRN 402  
Qy 367 RIGFAVSA 374  
Db 403 RVGFAEAA 410  
RESULT 11  
5217891-15  
Patent No. 5217891  
APPLICANT: BRAKE, ANTHONY J.;VAN DEN BERG, JOHAN A.  
TITLE OF INVENTION: DNA CONSTRUCTS CONTAINING A KLUYVEROMYCES  
A FACTOR LEADER SEQUENCE FOR DIRECTING SECRETION OF HETEROLOGOUS  
POLYPEPTIDES  
NUMBER OF SEQUENCES: 23  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/507,398  
FILING DATE: 09-APR-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 78,551  
FILING DATE: 28-JUL-1987  
SEQ ID NO:15:  
LENGTH: 458  
5217891-15

Query Match 13.0%; Score 279.5; DB 6; Length 458;  
Best Local Similarity 26.8%; Pred. No. 2.1e-21;  
Matches 95; Conservative 67; Mismatches 128; Indels 65; Gaps 15;  
Qy 31 YVEMTVGSPPTLNILVDTGSSNFAVGAAPHPFL-----HRYQRLSSYRDLKGV 83  
Db 152 YFGYLGTPQEFVTLFDTGSSDFWV---PSIYCKSNACKNHQRFQDKSSTFQNLGKDL 208  
Qy 84 YVPTQKWEGLTDLVSIPIHGPNTVVRANIAAITESDKFFINGSNWEGIILAYAEIA 143  
Db 209 SIHYGTGSMOGLGYDVTIVSNIVDTQOTVGLSTGEPGVF--TYAEFDGILGMAYPSLA 266  
Qy 144 RPDSLEPFDSLVKQTHV-PNLFSLQLCGAGFPLQSEVLASVGGSMIIGGDHSLYTG 202  
Db 267 --SEYISIPVEDNMNHLVAQDLFSVYMDRNG---QESMLT-----LGAIDPSYTG 313  
Qy 203 SLWYTPTRREWEYEVILVVEIINGODLKMD--CKEYNDKSIVDSTGNLRLPKKVFEAA 260  
Db 314 SLHWVPVTVQOYVQFTVDSVTISGVVYVACSGGC-----QAILDTGTSLKLVGFPSSDLNI 367  
Qy 261 VKSIKAASSTKFPDGFGLGE-QLVCWOAGTTPWNTFPVISLYLMEVTNQSPTITLPQ 319  
Db 368 QQAICATQNG-----YGEFDIDCDNLSYMPTVVF-----EINGKMYPLT--PS 408  
Qy 320 QYLRPVEDVATSDDCY---KFAISOSSGTVMGAVIMEGFYVVFDRARKRIGFA 371  
Db 409 AY-----TSQDQGFCTSGFQSENHSGWILGDVFIQRYTYVFDNRNVLGLA 455

RESULT 12  
US-09-640-305-6  
Sequence 6, Application US/09640305  
Patent No. RE37447  
GENERAL INFORMATION:  
APPLICANT: Fleer, Reinhard  
Fournier, Alain  
Yeh, Patrice  
TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Rd. 3c43  
CITY: Collegeville







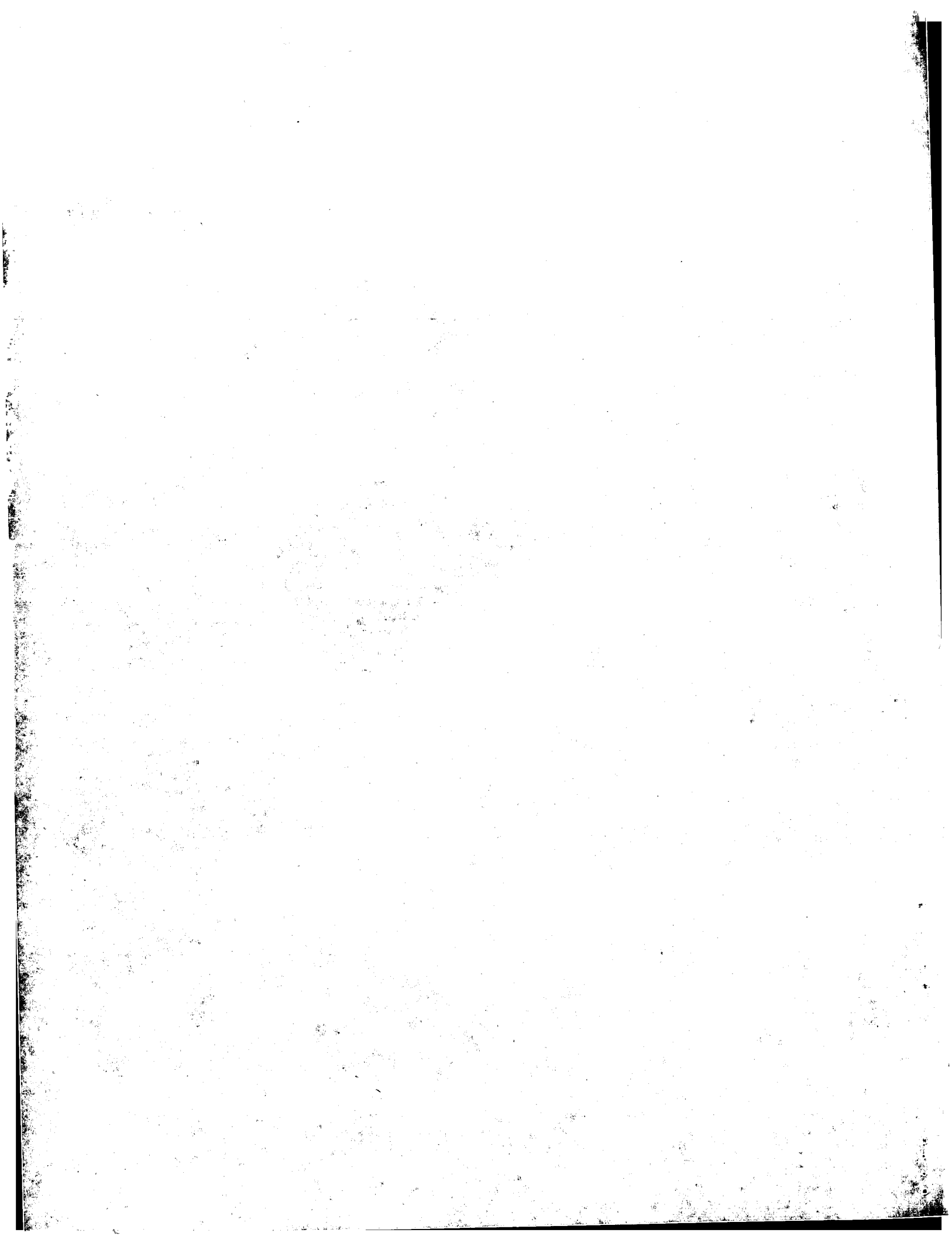
Wed Aug 7 10:38:05 2002

us-09-724-571-58.rai

Page 9

Db 357 DYTLEVSSCISAFPMDFP-EPIGPLAIIIGDSFLRKYYVYDLGKDAVGLAKS 409

Search completed: August 7, 2002, 09:14:49  
Job time: 74 sec



GenCore version 4.5  
 Copyright (c) 1993 - 2000 Compugen Ltd.  
 OM protein - protein search, using sw model  
 Run on: August 7, 2002, 09:13:35 ; Search time 56.55 Seconds  
 (without alignments)  
 895.661 Million cell updates/sec  
 Title: US-09-724-571-43  
 Perfect score: 2419  
 Sequence: 1 ETDEPPEPGRGSFVEMD.....CLRLRQHDDFADDISLLK 456  
 Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5  
 Searched: 747574 seqs, 111073796 residues  
 Total number of hits satisfying chosen parameters: 747574  
 Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000  
 Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2419	100.0	456	21	AA07897
2	2419	100.0	488	22	AA07897
3	2419	100.0	488	22	AA07897
4	2419	100.0	501	21	AA07897
5	2419	100.0	501	21	AA07897
6	2419	100.0	503	22	AA07897
7	2419	100.0	503	22	AA07897
8	2414	99.8	501	21	AA07897
9	2414	99.8	501	22	AA07897
10	2414	99.8	501	22	AA07897
11	2414	99.8	501	22	AA07897

12	2414	99.8	501	22	AAU07202	Human aspartyl pro
13	2414	99.8	501	22	AAE02581	Human aspartyl pro
14	2413	99.8	501	19	AAW59807	Amino acid sequenc
15	2392	98.9	501	21	AA074769	Rat beta-secretase
16	2390	98.8	501	21	AA074768	Murine beta-secret
17	2390	98.8	501	21	AA074768	Murine beta-secret
18	2390	98.8	501	22	AAE10631	Murine aspartyl pr
19	2390	98.8	501	22	AAE10631	Murine aspartyl pr
20	2390	98.8	501	22	AAE06861	Murine aspartyl pr
21	2390	98.8	501	22	AAU06605	Mouse aspartyl pr
22	2390	98.8	501	22	AAU07204	Mouse aspartyl pr
23	2390	98.8	501	22	AA074948	Mouse aspartyl sec
24	2351.5	97.2	969	22	AAE02583	Murine aspartyl pr
25	2320	95.9	790	19	AAW59808	Novel human diagno
26	2264.5	93.6	476	21	AA074768	Partial amino acid
27	2264.5	93.6	476	21	AA074768	Human aspartyl pro
28	2264.5	93.6	476	22	AAE10630	Human aspartyl pro
29	2264.5	93.6	476	22	AAE06860	Human aspartyl pro
30	2264.5	93.6	476	22	AAU06604	Human aspartyl pro
31	2264.5	93.6	476	22	AAU07203	Human aspartyl pro
32	2238.5	92.5	476	22	AAE02582	Human aspartyl pro
33	2238.5	92.5	476	22	AAE06909	Human aspartyl pro
34	2212	91.4	427	22	AA02619	Murine aspartyl pr
35	2160	89.3	453	21	AA073866	Murine aspartyl pr
36	2160	89.3	453	21	AA073866	Human polypeptide,
37	2160	89.3	453	22	AAE10642	Modified human asp
38	2160	89.3	453	22	AAE06872	Human-Asp 2(a) pro
39	2160	89.3	453	22	AAU06616	Human-Asp 2(a) del
40	2160	89.3	453	22	AAU07215	Human-pro-Asp 2(a)
41	2160	89.3	459	21	AA078439	Human aspartyl pro
42	2160	89.3	459	22	AAE10643	Modified human asp
43	2160	89.3	459	22	AAE10643	Human-Asp 2(a) pro
44	2160	89.3	459	22	AAU06617	Human-Asp 2(a) del
45	2160	89.3	459	22	AAU07216	Human-pro-Asp 2(a)
						Human aspartyl pro

## ALIGNMENTS

RESULT 1  
 AAB07897  
 ID AAB07897 standard; Protein; 456 AA.  
 XX AC  
 XX AAB07897;  
 XX DT 14-NOV-2000 (first entry)  
 XX DE Active enzyme portion of human beta-secretase enzyme.  
 XX KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
 KW inhibitor; ss.  
 XX OS Homo sapiens.  
 XX PN WO200047618-A2.  
 XX PD 17-AUG-2000.  
 XX PF 10-FEB-2000; 2000WO-US03819.  
 XX PR 10-FEB-1999; 99US-0119571.  
 PR 15-JUN-1999; 99US-0139172.  
 XX (ELAN-) ELAN PHARM INC.  
 XX PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
 PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
 XX WPI: 2000-533011/48.  
 PT Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.

```
PT Alzheimer's disease -
XX Claim 24; Fig 2B; 121pp; English.
XX
XX The specification describes a beta-secretase enzyme. The enzyme cleaves
XX beta-amyloid precursor protein to produce beta-amyloid peptide. This
XX enzyme is therefore implicated in the production of amyloid plaque
XX components which accumulate in the brains of individuals afflicted with
XX Alzheimer's disease. Inhibitors of beta-secretase are administered to
XX a mammalian subject e.g. with Alzheimer's disease or Alzheimer's
XX disease-like pathology to test if they maintain or improve cognitive
XX ability or reduce the plaque burden. The compounds are used for the
XX treatment of amyloidogenic diseases e.g. Alzheimer's disease. The
XX present sequence represents the active enzyme portion of human
XX beta-secretase enzyme.
XX Sequence 456 AA;
SQ
Query Match 100.0%; Score 2419; DB 21; Length 456;
Best Local Similarity 100.0%; Pred. No. 3.9e-242;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ETDEPEPEGRGSGFVEMVDNLKSGQGYVEMTVGSPPTNLILVDTGSSNFVAVGAAP 60
DB 1 etdeepeeprgrsgfvmvndnlrgskggyvemtvgspptnlilvdtgssnfavgaap 60
QY 61 HPFLHRYQRLSSTYRDLRKGVVYPTQGWEGELGTLVSIPIHGPNTVVRANIAITE 120
DB 61 hpflhryyqrqlsstyrdlrkgvvpytqgkwegelgtlvsiphgnvtvranaiaite 120
QY 121 SDKFFINGSNWEGILGLAYAEIARPDSDLEPFDLSLVKQTHVPLNLSLQLCGAGFPLNQS 180
DB 121 sdkffingsnwegilglayaeiarpdpsdlepfdslvkqthvplnlsfqlcgagfplnqs 180
QY 181 EYLASVGGSMIIGGDHSLYTGSLWYTPIRREWYVEVLIIVRVEINGQDLKMDCKEYNYDK 240
DB 181 evlasvggsmliggdhsltytgslytswytpirrewyvevliivrveingqdlkmdckeynydk 240
QY 241 SIVDSGTTNLRPKKVFEEAVKSIKAASSTKFPDGLVGCWQAGTTPWNIFFPVIS 300
DB 241 sivdsqgttnlrpkkvfeavksikaasstekfpdgglvqvcwagttppwnifpvvis 300
QY 301 LYLMGEVNTNQSPRITILPQQYLRPVEDVATSDDCYKFAISQSSSTGTVMGAVIMEGFYV 360
DB 301 lylmgevntnqsfritilpqylrpvedvatsddcykfaissqstgtvmgavimegfyyv 360
QY 361 FDRARRKRIGFAYSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDDESTLMTIAYVMAAI 420
DB 361 fdrarkrigfavsachvhdefrtaavegpfvtldmedcgynipqtdestlmtiayvmaai 420
QY 421 CALFMLPLCLMVCQWRCLRCLRQHQHDDFADDISLLK 456
DB 421 calfmplclmvcwqrclrclrqghddfaddisllk 456
RESULT 2
ID AAB66572 standard; Protein; 488 AA.
AC AAB66572;
XX
XX 12-APR-2001 (first entry)
DT
XX
XX Human memapsin 2.
DE
XX
XX Human; memapsin 2; nootropic; neuroprotective; amyloid precursor protein;
KW APP; memapsin 2 inhibitor; Alzheimer's disease.
XX
XX Homo sapiens.
OS
XX
XX WO200100665-A2.
PN
XX
```

```
PD 04-JAN-2001.
XX
XX 27-JUN-2000; 2000WO-US17742.
XX
XX 28-JUN-1999; 99US-0141363.
XX 30-NOV-1999; 99US-0168080.
XX 25-JAN-2000; 2000US-0177836.
XX 27-JAN-2000; 2000US-0178368.
XX 08-JUN-2000; 2000US-0210292.
XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.
XX (UNII ) UNIV ILLINOIS FOUND.
XX
XX Tang JJN, Hong L, Ghosh AK;
XX
XX WPI; 2001-137933/14.
XX N-PSDB; AAF31848.
XX
XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2
XX having 2 catalytic aspartic residues and substrate binding cleft, used
XX to treat Alzheimer's disease by blocking amyloid precursor protein
XX cleavage
XX
XX Example 1; Page 72-74; 86pp; English.
XX
XX The present sequence is given in a specification relating to an inhibitor
XX of catalytically active memapsin 2. The inhibitor binds to the memapsin 2
XX active site, which is defined by the presence of two catalytic aspartic
XX residues and a substrate binding cleft. The inhibitor is useful for
XX the treatment and diagnosis of Alzheimer's disease. It is useful in
XX screens for individuals with a genetic predisposition to Alzheimer's
XX disease. The inhibitor is useful as a reagent for specifically binding to
XX memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
XX isolation, purification and characterisation.
XX
XX Sequence 488 AA;
SQ
Query Match 100.0%; Score 2419; DB 22; Length 488;
Best Local Similarity 100.0%; Pred. No. 4.3e-242;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ETDEPEPEGRGSGFVEMVDNLKSGQGYVEMTVGSPPTNLILVDTGSSNFVAVGAAP 60
DB 33 etdeepeeprgrsgfvmvndnlrgskggyvemtvgspptnlilvdtgssnfavgaap 92
QY 61 HPFLHRYQRLSSTYRDLRKGVVYPTQGWEGELGTLVSIPIHGPNTVVRANIAITE 120
DB 93 hpflhryyqrqlsstyrdlrkgvvpytqgkwegelgtlvsiphgnvtvranaiaite 152
QY 121 SDKFFINGSNWEGILGLAYAEIARPDSDLEPFDLSLVKQTHVPLNLSLQLCGAGFPLNQS 180
DB 153 sdkffingsnwegilglayaeiarpdpsdlepfdslvkqthvplnlsfqlcgagfplnqs 212
QY 181 EYLASVGGSMIIGGDHSLYTGSLWYTPIRREWYVEVLIIVRVEINGQDLKMDCKEYNYDK 240
DB 213 evlasvggsmliggdhsltytgslytswytpirrewyvevliivrveingqdlkmdckeynydk 272
QY 241 SIVDSGTTNLRPKKVFEEAVKSIKAASSTKFPDGLVGCWQAGTTPWNIFFPVIS 300
DB 273 sivdsqgttnlrpkkvfeavksikaasstekfpdgglvqvcwagttppwnifpvvis 332
QY 301 LYLMGEVNTNQSPRITILPQQYLRPVEDVATSDDCYKFAISQSSSTGTVMGAVIMEGFYV 360
DB 333 lylmgevntnqsfritilpqylrpvedvatsddcykfaissqstgtvmgavimegfyyv 392
QY 361 FDRARRKRIGFAYSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDDESTLMTIAYVMAAI 420
DB 393 fdrarkrigfavsachvhdefrtaavegpfvtldmedcgynipqtdestlmtiayvmaai 452
QY 421 CALFMLPLCLMVCQWRCLRCLRQHQHDDFADDISLLK 456
DB 453 calfmplclmvcwqrclrclrqghddfaddisllk 488
```

RESULT 3  
ID AAB61334 standard; protein; 488 AA.  
XX  
AC AAB61334;  
XX  
DT 02-APR-2001 (first entry)  
XX  
DE Memapsin 2 protein.  
XX  
KW Memapsin 2; catalyst; Alzheimer's.  
XX  
OS Homo sapiens.  
XX  
PN WO200100663-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 27-JUN-2000; 2000WO-US17661.  
XX  
PR 28-JUN-1999; 99US-0141363.  
PR 30-NOV-1999; 99US-0168080.  
PR 25-JAN-2000; 2000US-0177836.  
PR 27-JAN-2000; 2000US-0178368.  
PR 08-JUN-2000; 2000US-0210292.  
XX  
PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
XX  
PI Tang JJN, Lin X, Koelsch G;  
XX  
DR WPI; 2001-102885/11.  
XX  
PT Purified recombinant catalytically active memapsin 2, used to screen  
PT inhibitors of it, which are used to treat and prevent Alzheimer's  
PT disease -  
XX  
PS Claim 2; Page 73-75; 86pp; English.  
XX  
CC The present invention relates to a purified recombinant  
CC catalytically active memapsin 2. The invention may be used for  
CC isolating inhibitors which are used to treat or prevent  
CC Alzheimer's disease. The invention may also be used to screen  
CC for individuals more genetically prone to develop Alzheimer's  
CC disease.  
XX  
SQ Sequence 488 AA;

Query Match 100.0%; Score 2419; DB 22; Length 488;  
Best Local Similarity 100.0%; Pred. No. 4.3e-242;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEPEGRGSEFVEMVNLKSGQGYVEMTVGSPQTLNILDVTGSSNFVGAAP 60  
Db 33 etdeepeepgrgsefvmvnlrgsggyvemtvgspqtlnilvdtgssnfavgap 92  
QY 61 HPFLHRYQRLSTYRDLRKGVPVPTQKWEGLCTDLVSPHGNVTVRANIAITE 120  
Db 93 hpflhryqrqlstydrlrgvvyptqkgwgelgtdlvsiphgnvtraniaaite 152  
QY 121 SDRFFINGSNWEGILGAYAEIARPDLSLFFDLSLVKQTHVNLFLSLQCGAGFPLNQS 180  
Db 153 sdrffingsnwegilgayaearpdslsldslvkvqthvnlflslqcgagfplnqs 212  
QY 181 EVLASVGSMLIGGIDHSILYTGSLWYPIRREYVYIIVRVEINGDLKMDCKEYNYDK 240  
Db 213 evlasvgssmliggidhslytgslywtpirreyvylivrvveingdldkmdckeynydk 272  
QY 241 SIYDVGTTNLRPKKVEAAVKSTKFAASSTKPPDGGFWLGEQLVCWAGTTPWNIFPVIS 300  
Db 273 siydvsgttnlrpkkveaavksikaasstekipdpgfwlgeqlvcwagttppwnifpvls 332

QY 301 LYLMEVTNQSFRTILPQQYLRPVEDVATSDDCYKFAISQSSGTGTVMGAVIMEGFYV 360  
Db 333 lylmevtnqsfritilppqylrpvedvatsqddcykfaissgtgtvmgavimegfyyv 392  
QY 361 FDRARKRIGFAVSACHVHDEFTAAVEGPFVTLDMEDCGYNIPOTDESTLMTIAVYVMAAI 420  
Db 393 fdrarkrigfavsachvhdefraavegpfvtldmedcgynlpqdestlmtiayvmaai 452  
QY 421 CALFMLPLCLMVCWMRCRLRCLRQHQHDDFADDISLIK 456  
Db 453 calfmlplcmvcwrcrlrclrqhghddfaddislik 488

RESULT 4  
AAI94767  
ID AAY94767 standard; Protein; 501 AA.  
XX  
AC AAY94767;  
XX  
DT 12-FEB-2001 (first entry)  
XX  
DE Human beta-secretase amino acid sequence.  
XX  
KW Beta-secretase; enzyme; amyloid plaque; Alzheimer's disease; human;  
KW Down's syndrome; amyloid angiopathy; gene therapy; neuroprotective.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..45  
FT /label= putative signal peptide  
FT Protein 46..501  
FT /label= Beta-secretase  
XX  
PN WO200058479-A1.  
XX  
PD 05-OCT-2000.  
XX  
PF 23-MAR-2000; 2000WO-US07755.  
XX  
PR 26-MAR-1999; 99US-0277229.  
XX  
PA (AMGE-) AMGEN INC.  
XX  
PI Citron M, Vassar RJ, Bennett BD;  
XX  
DR WPI; 2000-594643/56.  
DR N-PSDB; AAA28278.  
XX  
PT Isolated beta-secretase nucleic acids and encoded polypeptides, useful  
PT for diagnosis and gene therapy of Alzheimer's disease -  
XX  
PS Claim 1; Fig 4; 145pp; English.  
XX  
CC This invention relates to 3 nucleotide sequences encoding beta-secretase  
CC proteins. Beta-secretase is an enzyme involved in the production of one  
CC of the components of amyloid plaques involved in Alzheimer's disease. The  
CC invention includes an expression vector comprising the nucleotide  
CC sequence, a host cell comprising the expression vector, and a process for  
CC producing the protein through culturing the transformed cells. Also  
CC included in the invention are a polypeptide derivative of the  
CC beta-secretase protein, a fusion protein comprising beta-secretase fused  
CC to a heterologous amino acid sequence, and a method for modulating the  
CC levels of beta-secretase polypeptide in a mammal comprising administering  
CC the polynucleotide sequence. Beta-secretase exhibits neuroprotective and  
CC nootropic activity. The beta-secretase nucleotide sequence may be used to  
CC map locations of the beta-secretase gene and related genes on chromosomes  
CC and as hybridization probes in diagnostic assays to test for the presence  
CC of beta-secretase DNA or RNA, such as in Alzheimer's disease, Down's  
CC syndrome, and amyloid angiopathy. The nucleotide sequence may also be  
CC used as anti-sense inhibitors of beta-secretase expression, in gene  
CC therapy of Alzheimer's disease, and for the identification of compounds

CC that modulate beta-secretase activity. Antibodies to the beta-secretase  
CC protein may be used for in vitro and in vivo diagnostic purposes to  
CC detect the presence of beta-secretase polypeptide in a body fluid or cell  
CC sample. The present sequence represents the human beta-secretase protein.  
XX  
SQ Sequence 501 AA;

Query Match 100.0%; Score 2419; DB 21; Length 501;  
Best Local Similarity 100.0%; Pred. No. 4.5e-242;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ETDEEPEEGRGRSFVEMVNDLRGKSGQGYVEMTVGSPQTILNILDVTGSSNFAVCAAP 60  
Db 46 etdeepeepgrgrsfvemvndlrgrsggyvemtvgspqtnilildvtgssnfavgaap 105  
QY 61 HPFLHRYQRLSTYRDLRKGVVPTQGWEGELGTLVSIPIHGNVTVRANIAAITE 120  
Db 106 hpflhryyqrlstyrdlrgkvyvptqgkwegelgtlvsiphgpnvtvranaiaite 165  
QY 121 SDKFFINGSNWEGILGAYAEIARPDSDLVSKQTHVHVNLFSLQLCGAGFPLNQS 180  
Db 166 sdkffingsnwegilgayaearpddslvskqthvhnlfslqlcgagfplnqs 225  
QY 181 EVLASVGSMTLIGIDHSIYTLGSLWYPIRREWYVEVIIVRVEINGODLKMCKEYNVDK 240  
Db 226 evlasvgsmtligidhslytgslywypirrewyveviiivrveingodlkmckekyenydk 285  
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFWMGEQLVCMQAGTTPWNIFPVIS 300  
Db 286 sivdsgttnlrpkkvfeavksikaasstekfpdgfwlgeqlvcwqagttwnifpvvis 345  
QY 301 LYLMGEVTNOSFRITILPQQYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYVV 360  
Db 346 lylmgevtnosfrilitpqgylrpvedvatsddcykfaissqstgtvmgavimegyfvyv 405  
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420  
Db 406 fdrarkrigfavsachvhdefrtaavegpfvtldmedcgyinipqtdstlmtiayvmaai 465  
QY 421 CALFMLPLCLMVCOWRCLRLRQHQHDFADDISLLK 456  
Db 466 calfmlplclmvcowrcrlrlrqhqhdffaddisllk 501

RESULT 5  
AAB07896  
ID AAB07896 standard; Protein; 501 AA.  
XX  
AC AAB07896;  
XX  
DT 14-NOV-2000 (first entry)  
XX  
DE Amino acid sequence of a human beta-secretase enzyme.  
XX  
KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
KW inhibitor.  
XX  
OS Homo sapiens.  
XX  
FN WO200047618-A2.  
XX  
PD 17-AUG-2000.  
XX  
PF 10-FEB-2000; 2000WO-US03819.  
XX  
PR 10-FEB-1999; 99US-0119571.  
PR 15-JUN-1999; 99US-0139172.  
XX  
PA (ELAN-) ELAN PHARM INC.  
PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;

PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
XX  
DR WPI; 2000-533011/48.  
DR N-PSDB; AAA59550, AAA59551.  
XX  
XX Purified beta-secretase protein used in assays to discover inhibitors  
XX which can be used for the treatment of amyloidogenic diseases e.g.  
XX Alzheimer's disease  
PT  
PT  
XX  
XX Claim 17; Fig 2A; 121pp; English.  
PS  
XX  
CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
CC enzyme is therefore implicated in the production of amyloid plaque  
CC components which accumulate in the brains of individuals afflicted with  
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
CC disease-like pathology to test if they maintain or improve cognitive  
CC ability or reduce the plaque burden. The compounds are used for the  
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
CC present sequence represents a human beta-secretase enzyme.  
XX  
SQ Sequence 501 AA;

Query Match 100.0%; Score 2419; DB 21; Length 501;  
Best Local Similarity 100.0%; Pred. No. 4.5e-242;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ETDEEPEEGRGRSFVEMVNDLRGKSGQGYVEMTVGSPQTILNILDVTGSSNFAVCAAP 60  
Db 46 etdeepeepgrgrsfvemvndlrgrsggyvemtvgspqtnilildvtgssnfavgaap 105  
QY 61 HPFLHRYQRLSTYRDLRKGVVPTQGWEGELGTLVSIPIHGNVTVRANIAAITE 120  
Db 106 hpflhryyqrlstyrdlrgkvyvptqgkwegelgtlvsiphgpnvtvranaiaite 165  
QY 121 SDKFFINGSNWEGILGAYAEIARPDSDLVSKQTHVHVNLFSLQLCGAGFPLNQS 180  
Db 166 sdkffingsnwegilgayaearpddslvskqthvhnlfslqlcgagfplnqs 225  
QY 181 EVLASVGSMTLIGIDHSIYTLGSLWYPIRREWYVEVIIVRVEINGODLKMCKEYNVDK 240  
Db 226 evlasvgsmtligidhslytgslywypirrewyveviiivrveingodlkmckekyenydk 285  
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFWMGEQLVCMQAGTTPWNIFPVIS 300  
Db 286 sivdsgttnlrpkkvfeavksikaasstekfpdgfwlgeqlvcwqagttwnifpvvis 345  
QY 301 LYLMGEVTNOSFRITILPQQYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYVV 360  
Db 346 lylmgevtnosfrilitpqgylrpvedvatsddcykfaissqstgtvmgavimegyfvyv 405  
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420  
Db 406 fdrarkrigfavsachvhdefrtaavegpfvtldmedcgyinipqtdstlmtiayvmaai 465  
QY 421 CALFMLPLCLMVCOWRCLRLRQHQHDFADDISLLK 456  
Db 466 calfmlplclmvcowrcrlrlrqhqhdffaddisllk 501

RESULT 6  
AAB66573  
ID AAB66573 standard; Protein; 503 AA.  
XX  
AC AAB66573;  
XX  
DT 12-APR-2001 (first entry)  
XX  
DE Human pro-memapsin 2.  
XX  
KW Human; memapsin 2; nootropic; neuroprotective; amyloid precursor protein;



KW APP; memapsin 2 inhibitor; Alzheimer's disease; ss.

XX Homo sapiens.

XX WO200100665-A2.

XX 04-JAN-2001.

XX 27-JUN-2000; 2000WO-US17742.

XX 28-JUN-1999; 99US-0141363.

XX 30-NOV-1999; 99US-0168060.

XX 25-JAN-2000; 2000US-0177836.

XX 27-JAN-2000; 2000US-0178368.

XX 08-JUN-2000; 2000US-0210292.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX (UNII ) UNIV ILLINOIS FOUND.

XX Tang JJN, Hong L, Ghosh AK;

XX WPI; 2001-137933/14.

XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2

XX having 2 catalytic aspartic residues and substrate binding cleft, used

XX to treat Alzheimer's disease by blocking amyloid precursor protein

XX cleavage -

XX Example 4; Fig 1; 86pp; English.

XX The present sequence is given in a specification relating to an inhibitor

XX of catalytically active memapsin 2. The inhibitor binds to the memapsin 2

XX active site, which is defined by the presence of two catalytic aspartic

XX residues and a substrate binding cleft. The inhibitor is useful for

XX the treatment and diagnosis of Alzheimer's disease. It is useful in

XX screens for individuals with a genetic predisposition to Alzheimer's

XX disease. The inhibitor is useful as a reagent for specifically binding to

XX memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2

XX isolation, purification and characterisation.

XX Sequence 503 AA;

XX Query Match 100.0%; Score 2419; DB 22; Length 503;

XX Best Local Similarity 100.0%; Pred. No. 4.6e-242;

XX Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEEPEEPPGRGFSFVEMVDNLRKSGQGYVEMTVGSPQTNLILVDTGSSNFAVGAAP 60

DB 48 etdeepeepgrrgsfvemdnlrgsggyvemtvgspqptlnilvdtgssnfavgaa 107

QY 61 HPFLHRYYQRLSSTYRDLRKGVVVPYTGQKWGELGTLVSIPIHGNVTVRANIAAITE 120

DB 108 hpflhryyqrlstydrlrkgyvpytgqkwegeigtdivsphgnvtvranaiaite 167

QY 121 SDRFFINGSNWEGILGLAYAEIARPDPSLEPFFDSLKVQTHVFNLFSLQLCGAGFPLNQS 180

DB 168 sdrffingsnwegilglayaeiarpdpslepfdslnkvqthvfnlfsqlcgcagfplnqs 227

QY 181 EVLASVGSMTIIGIDHSILTYGSLWYPIRREWYEVIIIVRVEINGDGLKMDCKEYNYDK 240

DB 228 evlasvgssmli9gidhsiltygslwypirrewyeviiivrveingdglkmdckeynydk 287

QY 241 SIVDSGPTNLRLPKVFEAAVKSIKAASTKFPDGFGLGELVCMQAGTTPWNIFPVIS 300

DB 288 sivdsrgptnlrlpkvfeaaavksikaasstekfpdgfwlgeqlvcwaggttwnifpv 347

QY 301 LYLMEVTNSFRITILPQQYLRPVEDVATSQDDCYKFAISQSTGTVMGAVIMEGYVYV 360

DB 348 lylmgevtngsfrtilpqqylrpvedvatsqddcykfaissgstgtvmgavimegyf 407

QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420

DB 408 fdrarkrigfavsachvhdefrtaavegpfvltlmedcgynipqtdstlmtiayvmaai 467

QY 421 CALEWMLPCLIMVCQWRCLRLRQHQHDDFADDISILK 456

DB 468 calfmplcilmvcqwrclrlrqhnddfaddisilk 503

# RESULT 7

AAB61335

ID AAB61335 standard; protein; 503 AA.

XX AAB61335;

XX AC AAB61335;

XX 02-APR-2001 (first entry)

XX T7 promoter and vector sequence.

XX Memapsin 2; catalyst; Alzheimer's.

XX Homo sapiens.

XX Synthetic.

XX WO200100663-A2.

XX 04-JAN-2001.

XX 27-JUN-2000; 2000WO-US17661.

XX 28-JUN-1999; 99US-0141363.

XX 30-NOV-1999; 99US-0168060.

XX 25-JAN-2000; 2000US-0177836.

XX 27-JAN-2000; 2000US-0178368.

XX 08-JUN-2000; 2000US-0210292.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX Tang JJN, Lin X, Koelsch G;

XX WPI; 2001-102885/11.

XX Purified recombinant catalytically active memapsin 2, used to screen

XX inhibitors of it, which are used to treat and prevent Alzheimer's

XX disease -

XX Disclosure; Fig 1; 86pp; English.

XX The present invention relates to a purified recombinant

XX catalytically active memapsin 2. The invention may be used for

XX isolating inhibitors which are used to treat or prevent

XX Alzheimer's disease. The invention may also be used to screen

XX for individuals more genetically prone to develop Alzheimer's

XX disease.

XX Sequence 503 AA;

QY 1 ETDEEPEEPPGRGFSFVEMVDNLRKSGQGYVEMTVGSPQTNLILVDTGSSNFAVGAAP 60

DB 48 etdeepeepgrrgsfvemdnlrgsggyvemtvgspqptlnilvdtgssnfavgaa 107

QY 61 HPFLHRYYQRLSSTYRDLRKGVVVPYTGQKWGELGTLVSIPIHGNVTVRANIAAITE 120

DB 108 hpflhryyqrlstydrlrkgyvpytgqkwegeigtdivsphgnvtvranaiaite 167

QY 121 SDRFFINGSNWEGILGLAYAEIARPDPSLEPFFDSLKVQTHVFNLFSLQLCGAGFPLNQS 180

DB 168 sdrffingsnwegilglayaeiarpdpslepfdslnkvqthvfnlfsqlcgcagfplnqs 227

QY 181 EVLASVGSMTIIGIDHSILTYGSLWYPIRREWYEVIIIVRVEINGDGLKMDCKEYNYDK 240

DB 228 evlasvgssmli9gidhsiltygslwypirrewyeviiivrveingdglkmdckeynydk 287

Query Match 100.0%; Score 2419; DB 22; Length 503;  
Best Local Similarity 100.0%; Pred. No. 4.6e-242;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db 228 evlasvgsmiigdhsltytsglwtprirweyevliivrveingqdlkmdckeynydk 287
|||
QY 241 SIYDSGTTNLRPKKVFEEAAVKSIAKASSTKFPDGFWLGEQLVCWQAGTTPWNIFPVIS 300
|||
Db 288 sivdsgttnlrpkkvfeavksikaasstekfpgdfwlgelvcwgagttppwnifpvls 347
|||
QY 301 LYLMGEVTNQSFRITILPQOYLRPVEDVATSDQCYKFAISQSSTGTVMGAVIMEGFYVV 360
|||
Db 348 lyImgevtnsgsfrtitlppqyrlrpvedvatsgdcykfaIsqsgstgtvmgavImegfyvv 407
|||
QY 361 FDRARKRIGFAVSACHVHDFRFAAEGPFTVLDMEPCGYNIPQTDSTLMTIAYVMAAI 420
|||
Db 408 fdrarkrigfavsachvhdfraaevpgfvtldmedcgynipqtdestlmtiayvmaai 467
|||
QY 421 CALFMLEPLCLMVCQWRCLRCLRQHQHDDFADDISLLK 456
|||
Db 468 calfmplclmvcqwrclrcrlrgqhddfaddisilk 503
|||

RESULT 8
AA188425
ID AAY88425 standard; Protein; 501 AA.
XX
AC AAY88425;
XX
DT 03-AUG-2000 (first entry)
XX
DE Human aspartyl protease 2 (a) (Asp2) amino acid sequence.
XX
KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
KW Alzheimer's disease; beta secretase site.
XX
OS Homo sapiens.
XX
PN WO200017369-A2.
XX
PD 30-MAR-2000.
XX
PF 23-SEP-1999; 99WO-US20881.
XX
PR 24-SEP-1998; 98US-0101594.
XX
PA (PHAA ) PHARMACIA & UPJOHN CO.
XX
PI Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
XX
DR WPI; 2000-303209/26.
DR N-PSDB; AAA15662.
XX
PT New enzyme designated human aspartase useful in research into
PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at
PT the beta secretase site to produce amyloid beta peptide -
XX
PS Claim 48; Fig 2; 183pp; English.
XX
XX
This sequence represents the human aspartyl protease 2 (Asp2) amino acid
sequence. The invention relates to a protease (e.g. Asp2) capable of
cleaving the beta secretase site of amyloid precursor protein (APP). The
protease contains a sequence encoding the amino acid sequence DTG and a
sequence encoding DSG or DTG separated by 100-300 amino acids. When
mutated the APP gene causes an autosomal dominant form of Alzheimer's
disease. APP localises to the cell surface membrane and have a single
C-terminal transmembrane domain. Proteolytic processing of APP produces
the amyloid beta protein, which is possibly very important in Alzheimer's
disease. The invention includes a nucleotide sequence encoding the
protease, a vector containing the nucleotide sequence, and a cell line
comprising the vector. Methods for screening for inhibitors of beta
secretase activity are also given in the invention. The human aspartase
protein and nucleotide sequences and the methods for identifying
inhibitors of the protease, are useful in the treatment of and research
in to Alzheimer's disease.
XX
```

```
SQ Sequence 501 AA;
Query Match 99.8%; Score 2414; DB 21; Length 501;
Best Local Similarity 99.8%; Pred. No. 1.5e-241;
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDEEPEEPGRGSRFSFVEMVDNLRGKSGQGYVEMTGVSPQPTLNILVDTGSSFAVGAAP 60
|||
Db 46 etdeeepepprrgsfemvdnlrgksggyvemtvgspptlnilvdtgssnfavgaap 105
|||
QY 61 HPFLHRYQRQLSSTYRDLRKGVYVPYTOCKWEGELGTDLVSPHPGNVTVRANIAAITE 120
|||
Db 106 hpfLhryyqrqlsstyrDLrkgyvvytggkwegelgtdivsiphgnvtraniaaite 165
|||
QY 121 SDKFFFTNGSNWEGITGLAYAEIARPDSDLPEFFDSLVKQTHVPLNLSQLCGAGFPLNQS 180
|||
Db 166 sdkffingsnwegilglayaeiarpddsllepffdslvkqthvplnlshlcgagfplngs 225
|||
QY 181 EVLASVGSMIIGGIDHSLYTGSLWYTPIRREWYEVIIIVRVEINGODLKMCKEYNYDK 240
|||
Db 226 evlasvgsmiigdhsltytsglwtprirweyevliivrveingqdlkmdckeynydk 285
|||
QY 241 SIYDSGTTNLRPKKVFEEAAVKSIAKASSTKFPDGFWLGEQLVCWQAGTTPWNIFPVIS 300
|||
Db 286 sivdsgttnlrpkkvfeavksikaasstekfpgdfwlgelvcwgagttppwnifpvls 345
|||
QY 301 LYLMGEVTNQSFRITILPQOYLRPVEDVATSDQCYKFAISQSSTGTVMGAVIMEGFYVV 360
|||
Db 346 lyImgevtnsgsfrtitlppqyrlrpvedvatsgdcykfaIsqsgstgtvmgavImegfyvv 405
|||
QY 361 FDRARKRIGFAVSACHVHDFRFAAEGPFTVLDMEPCGYNIPQTDSTLMTIAYVMAAI 420
|||
Db 406 fdrarkrigfavsachvhdfraaevpgfvtldmedcgynipqtdestlmtiayvmaai 465
|||
QY 421 CALFMLEPLCLMVCQWRCLRCLRQHQHDDFADDISLLK 456
|||
Db 466 calfmplclmvcqwrclrcrlrgqhddfaddisilk 501
|||

RESULT 9
AAE10629
ID AAE10629 standard; Protein; 501 AA.
XX
AC AAE10629;
XX
DT 10-DEC-2001 (first entry)
XX
DE Human aspartyl protease 2(a) [hu-Asp2(a)] protein.
XX
KW Human; aspartyl protease 2(a); Asp2(a); amyloid precursor protein; APP;
KW Alzheimer's disease; AD; dementia; neurofibrillary tangle; gliosis;
KW amyloid plaque; neuronal loss; proteolytic; nontropic; neuroprotective;
KW chromosome 11q23.3-24.1.
XX
OS Homo sapiens.
XX
XX
Key Location/Qualifiers
XX Peptide 1..21
FT Peptide /label= Signal_peptide
FT Peptide 22..45
FT Peptide /label= Asp_2a_prepropeptide
FT Peptide 46..57
FT Peptide /label= Asp_2a_propeptide
FT Protein 58..501
FT Protein /label= Mature_human_Asp_2a_protein
FT Region 420..454
FT Region /label= Alpha-helical_spacer_region
FT Domain 455..477
FT Domain /label= Transmembrane_domain
FT Domain 478..501
FT Domain /label= Cytoplasmic_domain
XX
```



CC and neuronal loss. App isoforms are also used in methods for identifying  
CC inhibitors and modulators of human Asp2 activity. The invention relates  
CC to a method for identifying agents that modulate the activity of human  
CC aspartyl protease Asp2. Amyloid-beta peptides obtained from APP are used  
CC as a means to screen in cellular assays for the inhibitors of beta- and  
CC gamma- secretase. Hu-Asp DNA fragments are useful as probes or primers in  
CC polymerase chain reactions (PCR). The probes are useful for detecting  
CC Hu-Asp nucleic acids in in vitro assays and in Northern and Southern  
CC blots. The present sequence is human aspartyl protease 2 (Hu-Asp2), a  
CC 'long' form designated as (Hu-Asp2a). Hu-Asp 2 gene is localised on  
CC chromosome 11q23.3-24.1.  
XX  
SQ Sequence 501 AA;

Query Match 99.8%; Score 2414; DB 22; Length 501;  
Best Local Similarity 99.8%; Pred. No. 1.5e-241;  
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSGFVEMVDNLKSGGYVEMTVGSPPTNLILVDTGSSNFAVGAAP 60  
DB 46 etdeepeepgrgsfvmvmdnlrgsggyvemtvgspptnlilvdtgssnfavgaap 105  
QY 61 HPFLHRYQRLSSTYRDLRKGVVYPTQGWEGELGTLVSIPIHGPNTVVRANIAAITE 120  
DB 106 hpflhryyqrlsstyrdlrgvvyptqgkwegelgtlvsiphgpnvtvranaiaite 165  
QY 121 SDKFFINGNWEGILGLAYARIAPDPSLEFFDLSLVKQTHVPNLFSLQLCGAGFFLNQS 180  
DB 166 sdkffingsnwegilglayaeiarpdpsleffdsllvkqthvpnlfsllhcgagfflnqs 225  
QY 181 EVLASVGSMTIGDISHTYSLWYTPIRREWYEVILVIRVEINGDLMKDCKEYNVYDK 240  
DB 226 evlasvgsmtigdishtyqlwtpirrewyevilvrvveingqdlmkdckeynydk 285  
QY 241 SIVDSGTTNLRLPKKVFEEAAYKSIKAASSTBEPFDPGFWLGEQLVCWQAGTTPWIFPVIS 300  
DB 286 sivdsgttnlrlpkkvfeaaavksikaasstekfpgdglwlgelvcwqagttptwnifpv 345  
QY 301 LYLMEVTNSQFRITILPQOYLREVEDVATSDDCYKFAISOSSTGTVMGAVIMEGFVVV 360  
DB 346 lylmevtnsqfrtilpqoylrvedvatsddcykfaissgstgtvmgavimegfvvv 405  
QY 361 FDRARKRIGFAVSACHVHDEPRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420  
DB 406 fdrarkrigfavsachvhdeftaaavegpfvtlmedcgyinipqtdstlmtiayvmaai 465  
QY 421 CALFMLPCLMVCQWRCRLRCLRQHQHDFADDSILK 456  
DB 466 calfmplcmvcqwrcrlrcrqhqdffaddsilk 501

RESULT 11  
AAU06603  
ID AAU06603 standard; Protein; 501 AA.  
AC AAU06603;  
XX  
DT 24-OCT-2001 (first entry)  
XX  
DE Human Aspartyl protease 2(a), Asp2(a).  
XX  
KW Human; Aspartyl protease; Asp2(a); beta-secretase; neurotropic;  
KW neuroprotective; amyloid protein precursor; APP; Alzheimer's disease;  
KW amyloid-beta; Abeta.  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT Peptide /label= Signal peptide  
FT Peptide 22..45  
FT Peptide /label= Pre\_pro\_peptide

Peptide 46..57  
FT /label= Pro\_peptide  
FT Protein 57..501  
FT /label= Mature\_Asp2(a)  
FT Region 420..454  
FT Domian 455..477  
FT /label= Transmembrane\_domain  
FT Domian 478..501  
FT /label= Cytoplasmic\_domain  
XX  
PN WO200149098-A2.  
XX  
XX 12-JUL-2001.  
XX  
PD 09-MAY-2001; 2001WO-IB00798.  
PF  
XX 09-MAY-2001; 2001WO-IB00798.  
PR  
XX (BIEN/) BIENKOWSKI M J.  
PA (GURN/) GURNEY M E.  
PA (HEIN/) HEINRIKSON R L.  
PA (PARO/) PARODI L A.  
PA (YANR/) YAN R.  
XX  
PI Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;  
XX  
DR WPI: 2001-502549/55.  
DR N-PSDB; AAS11517.  
XX  
PT Novel purified polypeptide comprising fragment of mammalian aspartyl  
PT protease 2, lacking Asp2 transmembrane domain and retaining beta  
PT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
PT activity  
XX  
PS Claim 49; Fig 2; 185pp; English.  
XX  
CC The invention relates to a purified polypeptide comprising a fragment of  
CC mammalian aspartyl protease (Asp)2 protein which lacks the Asp2  
CC transmembrane domain and the Asp2 protein, and where the polypeptide and  
CC the fragment retain the beta-secretase activity of the mammalian Asp2  
CC protein. The invention also details polynucleotides for the Asp  
CC proteins and vectors expressing them, and a polypeptide (isoform of  
CC amyloid protein precursor (APP)) comprising the amino acid sequence of an  
CC APP or its fragment containing an APP cleavage site recognizable by a  
CC mammalian beta-secretase, and further comprising two lysine residues at  
CC the carboxyl terminus of the amino acid sequence of the mammalian APP or  
CC APP fragment. Also included in the invention are methods of identifying  
CC modulators or inhibitors of Asp2. Modulators and inhibitors of Asp2 are  
CC useful for treating Alzheimer's disease. APP is useful in methods for  
CC identifying inhibitors or modulators of human Asp2 activity and  
CC therapeutics for the treatment or prevention of Alzheimer's disease.  
CC APP comprising the APP-Sw-beta-secretase peptide sequence (NLDA), which  
CC is associated with increased levels of Abeta processing is useful in  
CC assays relating the Alzheimer's research. The expression vector is useful  
CC for recombinantly expressing APP. Nucleic acids that hybridise to  
CC Asp oligonucleotides are useful as probes or primers. The probes are  
CC useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
CC Northern and Southern blots. The present sequence is human Asp2(a).  
XX  
SQ Sequence 501 AA;

Query Match 99.8%; Score 2414; DB 22; Length 501;  
Best Local Similarity 99.8%; Pred. No. 1.5e-241;  
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSGFVEMVDNLKSGGYVEMTVGSPPTNLILVDTGSSNFAVGAAP 60  
DB 46 etdeepeepgrgsfvmvmdnlrgsggyvemtvgspptnlilvdtgssnfavgaap 105  
QY 61 HPFLHRYQRLSSTYRDLRKGVVYPTQGWEGELGTLVSIPIHGPNTVVRANIAAITE 120

Db 106 hpflhryyqrqlsystyrdlrgkgyvpytqgkwegelgtdivsiphgpnvtvranaiaite 165  
QY 121 SDKFFINGSNWEGTLGLAYAEIARPDLSLEPFDDSLVKQTHVPLNLSQLCGAGFPLNQS 180  
Db 166 sckffingsnwegilglayaeiarppdlslepffdsldvkqthvpnlfslhicgagfplnqs 225  
QY 181 EVLASVGSMIIIGIDHSLTGSWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNVDK 240  
Db 226 evlasvgsgmiigidhsltytgslywtpirrewyevliivrvveingqdlkmdckeynydk 285  
QY 241 SYVDSGTTNLRPKKVFEEAAVKSIKAASSTKEKPPDGFNLGELVLCWQAGTTPWNIFPVIS 300  
Db 286 sldvsgttnlrpkkvfeaaavksikaasstekfpdgfwlgeqlvcwqagttptwnifpvvis 345  
QY 301 LYLWGEVTNQSFRITILPQOYLPRVEDVATSDQCYKFAISQSSTGTVMGAVIMEGFYVV 360  
Db 346 lylmgevtnqsfritilpqgylrpvedvatsqdcykfaissgstgtvmgavimegfvyv 405  
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420  
Db 406 fdrarkrigfavsachvhdefrtaavegpfvtldmedcgyinipqtdestlmtiayvmaai 465  
QY 421 CALFMLEPLCLMVQWQRCRLRQHQHDFADDISLLK 456  
Db 466 calfmplclmvqcwrcrlrcrqghddfadddisllk 501

RESULT 12

AAU07202

ID AAU07202 standard; Protein: 501 AA.

XX AC AAU07202;

XX DT 24-OCT-2001 (first entry)

XX DE Human aspartyl protease 2a (Asp-2a).

XX KW Human; aspartyl protease 1; Asp-1; nootropic; neuroprotective;  
KW aspartyl protease 2; Asp2; amyloid protein precursor; APP;  
KW beta-secretase; Alzheimer's disease.  
XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..21

FT /note= "Signal peptide"

FT Misc\_feature 22..45

FT /note= "Pre-propeptide"

FT Misc\_feature 46..57

FT /note= "Propeptide"

FT Protein 58..501

FT /note= "Mature Aspartyl protease-2a"

FT Region 420..454

FT /note= "Alpha helical spacer region"

FT Domain 455-477

FT /note= "Transmembrane domain"

FT Domain 478..501

FT /note= "Cytoplasmic domain"

XX WO200149097-A2.

XX PD 12-JUL-2001.

XX PF 09-MAY-2001; 2001WO-IB00797.

XX PR 09-MAY-2001; 2001WO-IB00797.

XX PA (BIEN/) BIENKOWSKI M J.

XX PA (GURN/) GURNEY M E.

XX PA (HEIN/) HEINRIKSON R L.

XX PA (PARO/) PARODI L A.

XX PA (YANR/) YAN R.

XX PI Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;  
XX WPI: 2001-502548/55.  
XX N-PSDB: AAS11702.  
XX Novel purified polypeptide comprising fragment of mammalian aspartyl  
PT protease 2, lacking Asp2 transmembrane domain and retaining beta  
PT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
PT activity -

XX Claim 49; Fig 2: 185pp: English.

XX The invention relates to a novel purified polypeptide comprising a  
CC fragment of mammalian aspartyl protease 2 (Asp2) protein which lacks the  
CC Asp2 transmembrane domain and the Asp2 protein, and where the polypeptide  
CC and the fragment retain the beta-secretase activity of the mammalian Asp2  
CC protein. Also included is an isoform of amyloid protein precursor (APP)  
CC comprising the amino acid sequence of a APP or its fragment containing  
CC an APP cleavage site recognisable by a mammalian beta-secretase, and  
CC further comprising two lysine residues at the carboxyl terminus of the  
CC amino acid sequence of the mammalian APP or APP fragment. The  
CC polypeptides are used for assaying for modulators of beta-secretase  
CC activity; identifying agents that inhibit the APP processing activity  
CC of human Asp2 aspartyl protease (Hu-Asp2); identifying agents that  
CC modulate the activity of Asp2; and for reducing cellular production of  
CC amyloid beta (Abeta) from APP. Agents identified by the above methods  
CC are useful for treating Alzheimer's disease; and for identifying  
CC modulators of amyloid-beta (Abeta) peptide production, for use in  
CC designing therapeutics for the treatment or prevention of Alzheimer's  
CC disease. Probes and primers derived from Asp nucleic acid sequences  
CC are useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
CC Northern and Southern blots. The present sequence represents the  
CC amino acid sequence of human Asp-2a used in the methods of the invention.

XX SQ Sequence 501 AA;

Query Match 99.8%; Score 2414; DB 22; Length 501;  
Best Local Similarity 99.8%; Pred. No. 1.5e-241;  
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDEEPEEPGRGRGSFVEMVDNLRGKSGQGYVEMTVGSPPTQLNLYVDTGSSNFAVGAAP 60

Db 46 etdeepeepegrgrgsfvmvndnlrgksggyvemtvgspptqlnlyvdtgssnfavgaap 105

QY 61 HPFLHRYYQRLSSTYRDLRKGVVYVYPTQGWEGELGTDLVSIPIHGPNTVVRANIAITE 120

Db 106 hpflhryyqrqlsystyrdlrgkgyvpytqgkwegelgtdivsiphgpnvtvranaiaite 165

QY 121 SDKFFINGSNWEGTLGLAYAEIARPDLSLEPFDDSLVKQTHVPLNLSQLCGAGFPLNQS 180

Db 166 sckffingsnwegilglayaeiarppdlslepffdsldvkqthvpnlfslhicgagfplnqs 225

QY 181 EVLASVGSMIIIGIDHSLTGSWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNVDK 240

Db 226 evlasvgsgmiigidhsltytgslywtpirrewyevliivrvveingqdlkmdckeynydk 285

QY 241 SYVDSGTTNLRPKKVFEEAAVKSIKAASSTKEKPPDGFNLGELVLCWQAGTTPWNIFPVIS 300

Db 286 sldvsgttnlrpkkvfeaaavksikaasstekfpdgfwlgeqlvcwqagttptwnifpvvis 345

QY 301 LYLWGEVTNQSFRITILPQOYLPRVEDVATSDQCYKFAISQSSTGTVMGAVIMEGFYVV 360

Db 346 lylmgevtnqsfritilpqgylrpvedvatsqdcykfaissgstgtvmgavimegfvyv 405

QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420

Db 406 fdrarkrigfavsachvhdefrtaavegpfvtldmedcgyinipqtdestlmtiayvmaai 465

QY 421 CALFMLEPLCLMVQWQRCRLRQHQHDFADDISLLK 456

Db 466 calfmplclmvqcwrcrlrcrqghddfadddisllk 501

RESULT 13  
AAE02581  
ID AAE02581 standard; Protein; 501 AA.  
XX  
AC AAE02581;  
DT 10-AUG-2001 (first entry)  
XX  
DE Human aspartyl protease 2a (Asp 2a).  
XX  
KW Human; alpha-secretase; amyloid precursor protein; APP; therapy;  
KW Alzheimer's disease; antialzheimer's; aspartyl protease 2a; Asp 2a;  
KW beta-secretase; chromosome 11q23.3-24.1.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /label= Signal\_peptide  
FT Peptide 22..45  
FT /label= Asp\_2a\_prepropeptide  
FT Peptide 46..57  
FT /label= Asp\_2a\_propeptide  
FT Protein 58..501  
FT /label= Mature\_human\_Asp\_2a\_protein  
FT Active-site 93..95  
FT /label= Active\_site\_1  
FT Active-site 289..291  
FT /label= Active\_site\_2  
FT Region 420..454  
FT /label= Alpha\_helical\_spacer  
FT Domain 455..477  
FT /label= Transmembrane\_domain  
FT Domain 478..501  
FT /label= Cytoplasmic\_domain  
FT Region 486..501  
FT /note= "Peptide #2"  
XX  
PN WQ200123533-A2.  
XX  
PD 05-APR-2001.  
XX  
PF 22-SEP-2000; 2000WO-US26080.  
XX  
PR 23-SEP-1999; 99US-0155493.  
PR 23-SEP-1999; 99WO-US20881.  
PR 13-OCT-1999; 99US-0416901.  
PR 06-DEC-1999; 99US-0169232.  
XX  
PA (PHAA ) PHARMACIA & UPJOHN CO.  
XX  
PI Gurney M, Bienkowski MJ;  
XX  
DR WPI; 2001-290516/30.  
DR N-PSDB; AAD06739.  
XX  
PT Enzymes that cleave the alpha-secretase site of the amyloid precursor  
PT protein, useful for the treatment of Alzheimer's disease -  
XX  
PS Example 2; Fig 2; 189pp; English.  
XX  
CC The present invention relates to enzymes for cleaving the alpha-  
CC secretase site of the amyloid precursor protein (APP) and methods of  
CC identifying those enzymes. The methods may be used to identify enzymes  
CC that may be used to cleave the alpha-secretase cleavage site of the APP  
CC protein. The enzymes may be used to treat or modulate the progress of  
CC Alzheimer's disease. The present sequence is human aspartyl protease 2a  
CC (Asp 2a). Asp 2a has beta-secretase protease activity. Asp2 gene  
CC is located on chromosome 11q23.3-24.1.  
XX  
SQ Sequence 501 AA;

Query Match 99.8%; Score 2414; DB 22; Length 501;  
Best Local Similarity 99.8%; Pred. No. 1.5e-241;  
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ETDPEPEPEGRGRGSFVEMVDNLRGKSGQGYVEMTGVSPQTLNVLVDTSNPFANGAAP 60  
Db etdeepeepgrgrgsfvmvndlrgksqgyvemtgvspqtlnilvdtgssnfavgaap 105  
QY 61 HPFLHRYQRQLSSTYRDLRKGVVVPYTOCKWEGELGTLVSIPIHGPNVTVRANIAAITE 120  
Db hpfhlhryqrqlsstyrdlrkgvyvpytgkwegeigtldlvsiphgpnvtvranaiaaite 165  
QY 121 SDRFFFGNSWNEGILGLAYAEIARPDSDLPEFFDLSLVKQTHVPNLFSLQLCGAGFPLNQS 180  
Db sdrfffgnswnegilglayaeiarpdSDLPEFFDLSLVKQTHVPNLFSLQLCGAGFPLNQS 225  
QY 181 EVLASVGGSMIIGIDHSLYTGSLWYTPIRREWYVEVIVRVEINGQDLKMDCKEYNKYDK 240  
Db evlasvggsmiigidhsltytgslywtpirrewyvevivrveingqdlkmdckeynydk 285  
QY 241 SIYDSGTTLNLRPKKVFEEAAVKSIAASSTKPKPDGFMVLGEOLVCMQAGTTPWNIFPVIS 300  
Db siydsgttnlrpkkvfeaaavksikaasstekfpdgfmvlgeolvcmqagttwnifpvis 345  
QY 301 LYLMGEVTNQSFRITILPQQYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYV 360  
Db lylmgevtngsfrtilpqqylrpvedvatsqdqcykfaisqsgstgtvmgavimegfyyv 405  
QY 361 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTAYVMAAI 420  
Db fdrarrkrigfavsachvhdefrtaaavegpfvtldmedcgyNIPQTDSTLMTAYVMAAI 465  
QY 421 CALFMLPLCLMVQCWRCLRCLROQHDDFADDISLLK 456  
Db calfmplclmvqcwrcrlrclrqhddfaddisllk 501  
RESULT 14  
AAW59807  
ID AAW59807 standard; Protein; 501 AA.  
XX AC AAW59807;  
XX DT 26-OCT-1998 (first entry)  
XX DE Amino acid sequence of human ASP2 (aspartic protease 2).  
XX KW Human; ASP2; aspartic protease 2; agonist; antagonist; immunospecific;  
KW antibody; inhibition; Alzheimer's disease; cancer; proteinase;  
KW prohormone processing.  
XX OS Homo sapiens.  
XX PN EP855444-A2.  
XX PD 29-JUL-1998.  
XX PF 27-JAN-1998; 98EP-0300573.  
XX PR 28-JAN-1997; 97GB-0001684.  
XX (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
XX Chapman CG, Murphy K, Powell DJ, Smith TS;  
XX WPI; 1998-389809/34.  
DR N-PSDB; AAV41696.  
XX New nucleic acid encoding human aspartic protease 2 - used to treat,  
PT prevent and diagnose e.g. Alzheimer's disease, cancer and prohormone

```
PT processing
XX
PS Claim 1; Page 7; 26pp; English.
XX
CC This is the amino acid sequence of the human ASP2 (aspartic protease
CC family), used in the method of the invention. Agonists and
CC antagonists for ASP2 immunospecific antibodies are used to treat
CC conditions requiring increased or decreased activity or expression of
CC ASP2 respectively. ASP2 is used to treat and diagnose e.g.
CC Alzheimer's disease, cancer and pro-hormone processing and ASP2 or a
CC fragment can be used to induce an immune response against the above
CC conditions.
XX
SQ Sequence 501 AA;

Query Match 99.8%; Score 2413; DB 19; Length 501;
Best Local Similarity 99.8%; Pred. No. 1.9e-241;
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDEEPEEPPGRGSGFVEMVDNLRGKSGQGYVEMTVGSPPTNLTLVDTGSSNFVAGAAP 60
DB |||||||
QY 46 etdeepeepgrgsfvmvndlrgksgggyvemtvgppqtnlilvdtgssnfavgaap 105
DB |||||||
QY 61 HPFLHRYQRLSSTYRDLRGVYVPTQGWEGELGTDLSIPHGPNTVVRANIAAITE 120
DB |||||||
QY 106 hpflhryyqrqlsstyrdlrgkvypetqgkwegeigtldvslphgpnvtvranaiaite 165
DB |||||||
QY 121 SDKFFINGSNWEGILGLAYAEIARPDSDLEPFFDSLVKQTHVPNLFSQLCGAGFPLNQS 180
DB |||||||
QY 166 sdckffingsnwegilglayaeiarpdsllepffdslvkqthvplnflslqlcgagfplnqs 225
DB |||||||
QY 181 EVLASVGSMTIIGGDHSLYTGSLWYTPIRREWYVEIIVRVEINGODLKWDCKEYNYDK 240
DB |||||||
QY 226 evlasvgsmtiiggdhslsytgslwytpirrewyveiivrvveingodlkwmdckeynydk 285
DB |||||||
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLGQVQAGTTPWNIFPVIS 300
DB |||||||
QY 286 sivdsgttnlrpkkvfeaaavksiaasstekfpdgfwlgqlvcwaggttwnifpvvis 345
DB |||||||
QY 301 LYLMEVTNQSFRTITLPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYVV 360
DB |||||||
QY 346 lylmevtnqsfritilpqoylrpvedvatsddcykfaissgstgtvmgavimegfyyv 405
DB |||||||
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420
DB |||||||
QY 406 fdrarkrigfavsachvhdefrtaaavegpfvtdlmedcgynipqtdstlmtiayvmaai 465
DB |||||||
QY 421 CALFMLPLCLMVCWRCLRCLRQOHDDFADDISLLK 456
DB |||||||
QY 466 calfmplclmvcwrcrlrclrqohddfaddisllk 501
DB |||||||

RESULT 15
AAAY94769
ID AAY94769 standard; Protein; 501 AA.
XX
AC AAY94769;
XX
DT 12-FEB-2001 (first entry)
XX
DE Rat beta-secretase protein.
XX
KW Beta-secretase; enzyme; amyloid plaque; Alzheimer's disease; rat;
KW Down's syndrome; amyloid angiopathy; gene therapy; neuroprotective; ss.
XX
OS Rattus sp.
XX
PN WO200058479-A1.
XX
PD 05-OCT-2000.
XX
PF 23-MAR-2000; 2000WO-US07755.
```

```
XX
PR 26-MAR-1999; 99US-0277229.
XX
PA (AMGE-) AMGEN INC.
XX
PI Citron M, Vassar RJ, Bennett BD;
XX
DR WPT; 2000-594643/56.
DR N-PSDB; AAA28280.
XX
PT Isolated beta-secretase nucleic acids and encoded polypeptides, useful
XX for diagnosis and gene therapy of Alzheimer's disease -
XX
PS Claim 1; Fig 6; 145pp; English.
XX
CC This invention relates to 3 nucleotide sequences encoding beta-secretase
CC proteins. Beta-secretase is an enzyme involved in the production of one
CC of the components of amyloid plaques involved in Alzheimer's disease. The
CC invention includes an expression vector comprising the nucleotide
CC sequence, a host cell comprising the expression vector, and a process for
CC producing the protein through culturing the transformed cells. Also
CC included in the invention are a polypeptide derivative of the
CC beta-secretase protein, a fusion protein comprising beta-secretase fused
CC to a heterologous amino acid sequence, and a method for modulating the
CC levels of beta-secretase polypeptide in a mammal comprising administering
CC the polynucleotide sequence. Beta-secretase exhibits neuroprotective and
CC nontropic activity. The beta-secretase nucleotide sequence may be used to
CC map locations of the beta-secretase gene and related genes on chromosomes
CC and as hybridization probes in diagnostic assays to test for the presence
CC of beta-secretase DNA or RNA, such as in Alzheimer's disease, Down's
CC syndrome, and amyloid angiopathy. The nucleotide sequence may also be
CC used as anti-sense inhibitors of beta-secretase expression, in gene
CC therapy of Alzheimer's disease, and for the identification of compounds
CC that modulate beta-secretase activity. Antibodies to the beta-secretase
CC protein may be used in vitro and in vivo diagnostic purposes to
CC detect the presence of beta-secretase polypeptide in a body fluid or cell
CC sample. The present sequence represents the rat beta-secretase protein.
XX
SQ Sequence 501 AA;

Query Match 98.9%; Score 2392; DB 21; Length 501;
Best Local Similarity 98.2%; Pred. No. 2.9e-239;
Matches 448; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ETDEEPEEPPGRGSGFVEMVDNLRGKSGQGYVEMTVGSPPTNLTLVDTGSSNFVAGAAP 60
DB |||||||
QY 46 etdeepeepgrgsfvmvndlrgksgggyvemtvgppqtnlilvdtgssnfavgaap 105
DB |||||||
QY 61 HPFLHRYQRLSSTYRDLRGVYVPTQGWEGELGTDLSIPHGPNTVVRANIAAITE 120
DB |||||||
QY 106 hpflhryyqrqlsstyrdlrgkvypetqgkwegeigtldvslphgpnvtvranaiaite 165
DB |||||||
QY 121 SDKFFINGSNWEGILGLAYAEIARPDSDLEPFFDSLVKQTHVPNLFSQLCGAGFPLNQS 180
DB |||||||
QY 166 sdckffingsnwegilglayaeiarpdsllepffdslvkqthipnflslqlcgagfplnqt 225
DB |||||||
QY 181 EVLASVGSMTIIGGDHSLYTGSLWYTPIRREWYVEIIVRVEINGODLKWDCKEYNYDK 240
DB |||||||
QY 226 ealavsgsmiiggdhslsytgslwytpirrewyveiivrvveingodlkwmdckeynydk 285
DB |||||||
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLGQVQAGTTPWNIFPVIS 300
DB |||||||
QY 286 sivdsgttnlrpkkvfeaaavksiaasstekfpdgfwlgqlvcwaggttwnifpvvis 345
DB |||||||
QY 301 LYLMEVTNQSFRTITLPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYVV 360
DB |||||||
QY 346 lylmevtnqsfritilpqoylrpvedvatsddcykfaissgstgtvmgavimegfyyv 405
DB |||||||
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420
DB |||||||
QY 406 fdrarkrigfavsachvhdefrtaaavegpfvtdlmedcgynipqtdstlmtiayvmaai 465
DB |||||||
```

Accession	Protein	Score	Length	Ident	Positives	Negatives
Qy	421 CALFMLPLCLMVCQWRCLRCLRQHQHDDFADDISLLK	456	21	21	100%	0%
Db	466 calfmplclmvcqwrclrlrhghddfaddisllk	501	21	21	100%	0%

Search completed: August 7, 2002, 09:16:14  
Job time: 159 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:13:36 ; Search time 32.31 Seconds  
(without alignments)  
1356.135 Million cell updates/sec

Title: US-09-724-571-43  
Perfect score: 2419  
Sequence: 1 ETDEEPEEPRGRGSFVEMVD.....CLRLRQHQHDDFADDSILLK 456  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2419	100.0	501	2 A59090	aspartic proteinase
2	308	12.7	384	2 JC7574	pepsinogen A - Afr
3	308	12.7	385	2 JC7575	pepsinogen A - bul
4	306	12.6	387	2 B38302	pepsin (EC 3.4.23)
5	303	12.5	383	2 JC7573	pepsinogen C - Afr
6	302.5	12.5	388	1 S19082	pepsin A (EC 3.4.2
7	301	12.4	382	1 PECH	pepsin A (EC 3.4.2
8	299.5	12.4	396	2 A34401	cathepsin E (EC 3
9	298.5	12.3	383	2 A41443	pepsin (EC 3.4.23)
10	298.5	12.3	384	2 A39314	gastricsin (EC 3.4
11	296	12.2	387	2 C38302	pepsin (EC 3.4.23)
12	296	12.2	391	2 A43356	cathepsin E (EC 3
13	295.5	12.2	412	1 KHHUD	cathepsin D (EC 3
14	295	12.2	387	2 D38302	pepsin (EC 3.4.23)
15	289.5	12.0	388	1 S19684	pepsin (EC 3.4.23)
16	287.5	11.9	444	2 T24204	hypothetical prote
17	287	11.9	407	1 KHRD	cathepsin D (EC 3
18	286	11.8	387	2 E38302	pepsin (EC 3.4.23)
19	285.5	11.8	398	2 S66465	cathepsin E (EC 3
20	285	11.8	398	2 S11185	cathepsin D (EC 3
21	283.5	11.7	388	1 PEMQAR	pepsin A (EC 3.4.2
22	282.5	11.7	388	1 PEHU	pepsin A (EC 3.4.2
23	281.5	11.6	381	1 CMSEB	chymosin (EC 3.4.2
24	281.5	11.6	388	1 PEMQAJ	pepsin A (EC 3.4.2
25	281.5	11.6	388	2 A30142	pepsin A (EC 3.4.2
26	279.5	11.6	388	2 B30142	pepsin A (EC 3.4.2
27	279.5	11.6	410	1 KHMSD	cathepsin D (EC 3
28	278.5	11.5	386	1 PEPC	pepsin A (EC 3.4.2
29	278	11.5	387	2 JC7245	pepsinogen A - com

30	277.5	11.5	380	2 I47176	chymosin (EC 3.4.2
31	277.5	11.5	396	2 S36865	cathepsin E (EC 3
32	276	11.4	389	2 JEO371	pepsin C (EC 3.4.2
33	273.5	11.3	381	1 CMBO	chymosin (EC 3.4.2
34	270.5	11.2	377	1 PEMOCJ	gastricsin (EC 3.4
35	270.5	11.2	389	2 A38302	pepsin (EC 3.4.23)
36	270	11.2	376	2 I45856	aspartic proteinase
37	268.5	11.1	344	1 KHPGD	cathepsin D (EC 3
38	267.5	11.1	381	2 JC7247	prochymosin - comm
39	266	11.0	380	2 S03433	candidapepsin (EC
40	266	11.0	405	2 A25379	saccharopepsin (EC
41	264	10.9	396	2 T47207	aspartic proteinase
42	263.5	10.9	388	2 JC7246	pepsinogen C - com
43	262.5	10.9	394	2 B43356	gastricsin (EC 3.4
44	261.5	10.8	387	2 A45117	aspartic proteinase
45	261.5	10.8	388	2 A29937	gastricsin (EC 3.4

ALIGNMENTS

RESULT 1  
A59090  
aspartic proteinase (EC 3.4.23.-) BACE precursor - human  
N:Alternate names: beta-secretase, beta-site APP cleaving enzyme  
C:Species: Homo sapiens (man)  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 11-May-2000  
C:Accession: A59090  
R:Vassar, R.; Bennett, B.D.; Babu-Khan, S.; Kahn, S.; Mendiaz, E.A.; Denis, P.; Teplo  
M.A.; Biere, A.L.; Curran, E.; Burgess, T.; Louis, J.C.; Collins, F.; Treanor, J.; Ro  
Science 286, 735-741, 1999  
A:title: beta-Secretase cleavage of Alzheimer's amyloid precursor protein by the tran  
A:Reference number: A59090; MUID:2002972  
A>Note: submitted to GenBank, September 1999  
A:Accession: A59090  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-501 <VAS>  
A:CROSS-references: GB:AF190725; NID:g6118538; PIDN:AAF04142.1; PID:g6118539  
C:Genetics:  
C:Gene: BACE  
C:Superfamily: beta-secretase  
C:Keywords: Alzheimer's disease; aspartic proteinase; brain; glycoprotein; hydrolase;  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-45/Domain: propeptide #status predicted <PRO>  
F:46-501/Product: acid proteinase BACE #status predicted <MAT>  
F:461-477/Domain: transmembrane #status predicted <TRN>  
F:93,289/Active site: Asp #status predicted  
F:153,172,223,354/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:330-380/Disulfide bonds: #status predicted

Query Match 100.0%; Score 2419; DB 2; Length 501;  
Best Local Similarity 100.0%; Pred. No. 3.9e-192;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ETDEEPEEPRGRGSFVEMVDNLGRKSGQYGYVEMTVGSPQTLNIIIVDTGSSNFAYGAAP	60
DB	46	ETDEEPEEPRGRGSFVEMVDNLGRKSGQYGYVEMTVGSPQTLNIIIVDTGSSNFAYGAAP	105
QY	61	HPFLHRYQRQLSSTYRDLRKGVVYPYTGCKWEGELGTLVSTPHGPNVTVRANTAAITE	120
DB	106	HPFLHRYQRQLSSTYRDLRKGVVYPYTGCKWEGELGTLVSTPHGPNVTVRANTAAITE	165
QY	121	SKFFINGNWEIGILGLAYAEIARPDSDLVQTHVFNLFSLQCGAGPPLNS	180
DB	166	SKFFINGNWEIGILGLAYAEIARPDSDLVQTHVFNLFSLQCGAGPPLNS	225
QY	181	EVLASVGGSMIIGGDHSLYTGSLWTPTRREWYVEIIVRVEINGQDLKMDCKEYNDK	240
DB	226	EVLASVGGSMIIGGDHSLYTGSLWTPTRREWYVEIIVRVEINGQDLKMDCKEYNDK	285
QY	241	SIIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFNLGQVLCVQAGTTPWNIFPVIS	300

```
|||||
Db 286 SIYDSGTTNLRPKKFEAAVSIKAASTEKEFPDGFWLGEOQLVCHQAGTTPNIFPVLS 345
QY 301 LYLMEVNTGNSFRITILPQOYLRPVEDVATSDQDCYKFAISQSTGTVMGAVIMEGFYVY 360
Db 346 LYLMEVNTGNSFRITILPQOYLRPVEDVATSDQDCYKFAISQSTGTVMGAVIMEGFYVY 405
QY 361 FDRARKRIGFAVACHVHDFRTAAVEGPPVTILDMEDCGYNIPQTDSTLMTIAYVMAAI 420
Db 406 FDRARKRIGFAVACHVHDFRTAAVEGPPVTILDMEDCGYNIPQTDSTLMTIAYVMAAI 465
QY 421 CALFMPLCLMLVCMQWRCRLRQHQHDDFADDISILK 456
Db 466 CALFMPLCLMLVCMQWRCRLRQHQHDDFADDISILK 501

RESULT 2
JC7574
pepsinogen A - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
C:Accession: JC7574; PC7119
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens
A:Reference number: JC7573; MUID:21064922; PMID:11134969
A:Contents: Stomach
A:Accession: JC7574
A:Molecule type: mRNA
A:Residues: 1-384 <IKU>
A:Cross-references: DDBJ:AB045380
A:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like
A:Gene: PgA
C:Superfamily: pepsin
C:Keywords: stomach; zymogen

Query Match 12.7%; Score 308; DB 2; Length 384;
Best Local Similarity 27.5%; Pred. No. 1.3e-17;
Matches 99; Conservative 60; Mismatches 133; Indels 68; Gaps 16;
QY 30 YVEMTVGSPPTLNILVDGTSSNFVAVGAPHPFL-----HRYQQLSSTYRDLRKG 82
Db 72 YGTISIGTPPQSFVIFDGTGSAWLW---PSVYCSSQACSNHNRFPQOQSTFQATNP 128
QY 83 VVVPYTGKWEGLGTDLYSIHPGNVTVRANIAAITESDK--FFINGSNWEGILGLAYAE 141
Db 129 VSIQYTGSMGFLGYDPLQV---GNIQISNQMFGLSEPEGSFLYSPFDGILGLAFPS 185
QY 142 IARPDSDLPEFDSLVKQTHVP--NLFSLOLCGAGFPLNQSEVLASVGSMTIIGDHSLY 200
Db 186 TA--SSQATPVFDNMWSOGLIPQNLFSVYSSDG-----QTGSYVIFGGVDNSYY 233
QY 201 TGLWYTPIRREYEVIIIVRVEINGQDL--KMDCKEYNDKSIDVSGTTLNLRPKKVF 258
Db 234 SGLSNWPLTAETWQITLDSVINGQVIACQSC-----QAIVDTGTSMTGPTPI- 286
QY 259 AAVKSIKAASSTEKFPDGFWLGEOQLV--CWQAGTTPWNIFPVISLYLMGEVNTQSFRTIL 317
Db 287 ANIQNYIGASQDSN-----GQVYINCNINSMPTIVF-----TIN 321
QY 318 PQOY--LRPVEDVATSDQCYK--FAISQSSTGT---VMGAVIMEGFYVVFDRARKRIGFA 371
Db 322 GVQYPLSPSAVYRQNOQCGSSGFQAMNLTNSGDLWLIGDVFIRQYTFVFDNRANNVIA 381

RESULT 3
JC7575
pepsinogen A - bullfrog
C:Species: Rana catesbeiana (bullfrog)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
C:Accession: JC7575
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens
A:Reference number: JC7573; MUID:21064922; PMID:11134969
A:Contents: Stomach
A:Accession: JC7575
A:Molecule type: mRNA
A:Residues: 1-385 <IKU>
A:Cross-references: DDBJ:AB045376
A:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like
A:Gene: PgA
C:Superfamily: pepsin
C:Keywords: stomach; zymogen

Query Match 12.7%; Score 308; DB 2; Length 385;
Best Local Similarity 27.8%; Pred. No. 1.3e-17;
Matches 99; Conservative 64; Mismatches 133; Indels 60; Gaps 15;
QY 30 YVEMTVGSPPTLNILVDGTSSNFVAVGAPHPFLHRYYQQLSSTYRDLRKGVV 85
Db 73 YFTISIGTPPQSFVIFDGTGSSNLWVPSVYCSPACTNHHMFNPQOQSTFQATNPVSI 132
QY 86 PYTGKWEGLGTDLYSIHPGNVTVRANIAAITESDK--FFINGSNWEGILGLAYAEIAR 144
Db 133 QYGTGSMGFLGYDTPVQV---GNIQITNQIFGLSQSEPGSFLYSPFDGILGLAFPSLA- 188
QY 145 PDDSLPEFDSLVKQTHVP--NLFSLOLCGAGFPLNQSEVLASVGSMTIIGDHSLYTGS 203
Db 189 -SSQATPVFDNMWNOGLIPQDLFSVYSSQG-----GSFVFGVDTSYTG 237
QY 204 LWTPTIRREYEVIIIVRVEINGQDLK--DCKEYNDKSIDVSGTTLNLRPKKVFEEAV 261
Db 238 LNWVPLTAETWQITLDSVINGQVIACQSC-----SAIVDTGTSLAGP----STPI 287
QY 262 KSIKAASSTEKFPDGFWLGEOQLVCHQAGTTPWNIFPVISLYLMGEVNTQSFRTILPQY 321
Db 288 ANIQYIGANQDSNGQYV---INCNNISNMPTVVF-----TINGVQY 326
QY 322 LRPVED--VATSDQDC---YKFAISQSSTGT--VMGAVIMEGFYVVFDRARKRIGFA 371
Db 327 PLPASAYVROSQDSCTSGFQAMNLTSSGDLWLIGDVFIREYTVVFDNRANNVAMA 382

RESULT 4
B38302
pepsin (EC 3.4.23.-) II-1 precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 23-Feb-1997
C:Accession: B38302
R:Kageyama, T.; Tanabe, K.; Koiwai, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu
A:Reference number: A38302; MUID:91009127
A:Accession: B38302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:M59235; GB:J05638
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 12.6%; Score 306; DB 2; Length 387;
Best Local Similarity 27.1%; Pred. No. 2e-17;
Matches 98; Conservative 67; Mismatches 131; Indels 66; Gaps 15;
QY 30 YVEMTVGSPPTLNILVDGTSSNFVAVGAPHPFLHRYYQQLSSTYRDLRKGVV 85
Db 73 YFTISIGTPPQSFVIFDGTGSSNLWVPSVYCSPACTNHHMFNPQOQSTFQATNPVSI 132
```

Db 75 YFTGISIGPPQEFVIFDTGSSNLWVPSTYSSSLACFLHKRFNPDDSSFTQATSETLSI 134  
QY 86 PYTQKWECELTDLVSIPIHGNVTVRANIAALTESD---KEFINGSNNEGILGLAYAEI 142  
Db 135 TYGTGSMTCILGYDIVKV---GNIEDTNGIFGLSKTEPITFLV--APDGGILGLAYPSI 189  
QY 143 ARPDSLEPFPSLVKQTHV-PNLFSLQLCGAGFPNQLSEVLASVGGSMIIGDHSLYT 201  
Db 190 SASDAT--PVFDNMWNEGLVSEDLFSVYLSNG-----EKGSMVMFGGIDSSYYT 237  
QY 202 GSLWTPTRREWYEVIIIVRVEINGQDLKM--DCKEYNVDKSIDVSGTINLRLPKVFEA 259  
Db 238 GSLNWPVSHGTYWQITMDISITINGIACADSC-----QAVDVTGTLGAGPTSATSK 291  
QY 260 AVKSIKAASSTKFPDGFGEWLGQLV-CWQAGTTPNIFPVISLYLMGEVTNQSFRTITLP 318  
Db 292 IQSYIGASKNL-----LGENIISCAIDSLPDIVF-----TINN 325  
QY 319 QOYLPRVED-VATSDDC---YKFAISQSSTGT--VMGAVIMEGYVYVDFDRARKRIGFAV 372  
Db 326 VOYPLPASAYILKEDDCLSGFDGMNLDTSYGELWILGDVFIHQYFTVDFDRANNQVGLAA 385  
QY 373 SA 374  
Db 386 AA 387

RESULT 5  
JC7573  
pepsinogen C - African clawed frog  
N:Alternate names: progastricsin  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 03-Aug-2001  
C:Accession: JC7573; PC7118  
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.  
J. Biochem. 129, 147-153, 2001  
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens  
A:Reference number: JC7573; MUID:21064922; PMID:11134969  
A:Contents: Stomach  
A:Accession: JC7573  
A:Molecule type: mRNA  
A:Residues: 1-383 <IKU>  
A:Cross-references: DBJ:AB045379  
A:Accession: PC7118  
A:Molecule type: protein  
A:Residues: 17-68 <IK2>  
C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like a  
C:Genetics:  
A:Gene: Pgc  
C:Superfamily: pepsin  
C:Keywords: stomach; zymogen

Query Match 12.5%; Score 303; DB 2; Length 383;  
Best Local Similarity 27.4%; Pred. No. 3.4e-17;  
Matches 104; Conservative 57; Mismatches 122; Indels 96; Gaps 17;  
QY 30 YVEMTVGSPQTLNILDVDTGSSNFAVGA-----APHPFLHRYQRLSSTYDLRK 81  
Db 67 YGEISICTPQNFVLVDFDTGSSNLWASTVCQSQACTNHPL----FNPSQSTYSNQ 122  
QY 82 GYVPYPTQKWECELTDLVSIPIHGNVTVRANIAALTESDCKFFINGSN-----WEGILG 136  
Db 123 QFSLOYGTGSLTGILGYDTVTI---QNVAISQEQEGLSETEP---GTFNFAVQFDGILG 175  
QY 137 LAYAEIAPDDSLPEFPFSLVVKQTHVFNLFSLQLCGAGFPNQLSEVLASVGGSMIIGID 196  
Db 176 LAYPSIA--VGATTVMGMMQ-----NLLNQPI--FGYLSGQS--SONGGEVAFGGVD 225  
QY 197 HSLTSGSLWYTPIRREWYEVIIIVRVEINGOD---LKMCKEYNVDKSIDVSGTINLRLP 253  
Db 226 QNYITGYIWPVTSETWQIGIQFSINGATGWCSGC-----QAIVDTGTSLLTAP 279

QY 254 KVFEEAAVKSIAASSTKFPDGFGEWLGQLVCMQAGTTPWNI--PPVISLYLMG----- 305  
Db 280 QSVFSLSIGAQODQN-----GQVVSQS-----NIQNLTPTISFTTISGVFPPLP 325  
QY 306 ---EVTNQS-----FRITILPQOYLPRVEDVATSQDDCYKFAISQSSTGTVMGAVIME 355  
Db 326 PSAYVLQSSGYCTIGIMPTYLPSONGQPL-----WILGDVFLR 364  
QY 356 GFYVVFDRARKRIGRAVSA 374  
Db 365 EYISYVDLGNNOVGFEATAA 383  
RESULT 6  
S19682  
pepsin A (EC 3.4.23.1) 4 precursor - Japanese macaque  
N:Alternate names: pepsinogen A isozyme 4  
C:Species: Macaca fuscata (Japanese macaque)  
C:Date: 22-Nov-1993 #sequence\_revision 19-Oct-1995 #text\_change 18-Jun-1999  
C:Accession: S19682; S16065  
R:Kageyama, T.; Tanabe, K.; Koiwai, O.  
Eur. J. Biochem. 202, 205-215, 1991  
A:Title: Development-dependent expression of isozymogens of monkey pepsinogens and st  
A:Reference number: S19681; MUID:92037645  
A:Accession: S19682  
A:Molecule type: mRNA  
A:Residues: 1-388 <KAG>  
A:Cross-references: EMBL:X59753; NID:g38070; PIDN:CAA42425.1; PID:g38071  
C:Note: parts of sequence, including amino ends of pepsinogen and activation intermed  
C:Comment: This is a minor component of pepsin at all post-partum stages.  
C:Comment: Although two-step activation is observed, activation is predominantly a o  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein di  
F:1-15/Domain: signal sequence #status predicted <SIG>  
F:16-388/Product: pepsinogen A #status experimental <PPT>  
F:16-62/Domain: activation peptide #status experimental <APT>  
F:63-388/Product: pepsin A #status experimental <BNZ>  
F:38-39/Cleavage site: Leu-Lys (pepsin) #status experimental  
F:62-63/Cleavage site: Leu-Ile (pepsin) #status experimental  
F:94-277/Active site: Asp #status predicted  
F:107-112,268-272,311-344/Disulfide bonds: #status predicted  
F:130/Binding site: phosphate (Ser) (covalent) #status predicted  
Query Match 12.5%; Score 302.5; DB 1; Length 388;  
Best Local Similarity 27.6%; Pred. No. 3.8e-17;  
Matches 107; Conservative 64; Mismatches 133; Indels 83; Gaps 17;  
QY 3 DEEPEPGRGRSFVEMVDNLRGSGQGYVEMTVGSPQTLNILDVDTGSSNFAVGAAPHP 62  
Db 64 DEQPLE----NYLDV-----EYFGTIGTGAQNFVDFDTGSSNLWV---PSV 105  
QY 63 FL-----HRYQRLSSTYRDLRKGVVYPTQKWECELTDLVSIPIHGNVTVRANI 115  
Db 106 YCYSLACMDHNLFPQDSSTYRATSKTVSITYGTGSMTCILGYDTVKV---GGISDTNOI 162  
QY 116 RAITESDK-FFTINGSNWEGLGLAYAEIARPDSDLPEPFDLSLVKQTHV-PNLFSLQLCGA 173  
Db 163 FGLSTEGFFLYFPAPFDGILGLAYPSIS--SSGATPFDNLWNLQRLVSDQFVSLSD 220  
QY 174 GFPLNQSEVLASVGGSMIIGDHSLYTGLWYTPTRREWYEVIIIVRVEINGQDL--KM 231  
Db 221 ---DQS-----GSVWIFGGIDSSYTGSLNWPVSVGEYVQWISVDSITMNGKIACAK 270  
QY 232 DCKEYNVDKSIDVSGTINLRLPKVFEAAVKSIAASSTKFPDGFGEQLV-CWQAGT 290  
Db 271 GC-----QAIVDTGTSLLTGTPTSPITANIQSDIGASENSD-----GEMVVSCTAIS 316  
QY 291 TPWNITPVISLYLMGEVTNQSFRTILPQOY--LRPVEDVATSQDDCYK-----FAISQS 344  
Db 317 LPDIVF-----TINGVQYPLPPPSAYILQSGSGCTSGFGQGMVPTESG\_358  
QY 345 TGTVMGAVIMEGYVYVDFDRARKRIGFA 371



Db 170 TEPGQTEVD-AEFDGILGLGYPSLA--VGGVTPVFDNNMAQ-----NLVDLPMSFVSMSSN 222  
QY 179 QSEVLASVGGSMIIIGGDDHSLYTGSLAWTPIRREWYVEIIVRVEINGQDLKMDCKEYNY 238  
Db 223 PE---GGAGSELIFGGDHRHSGSLNWPVTKQAYQIALDNIOVGG--TVMFCSE--G 275  
QY 239 DKSIVDSGTTNLRPLKPKVFAAASIKAAASSTKFPDGFGLGEOLVCWQAGTTPTWNIFFPV 298  
Db 276 CQAIIVDTGTSITGSPDKIKOLQNAICAAP-----VDGEYAVE-----CANLNVMPD 322  
QY 299 ISLYLMGEVTVNQSPRIITLQQYLRLPVEDVATSDQDCYKFAISSSTG----- 346  
Db 323 VTFTING-----VPYTLSPAY--TLIDFVDMGMQFC-----SSGFGGLDIHPPAGP 366  
QY 347 -TVMGAVIMEGFYVVDRAKRKIGFA 371  
Db 367 LWILGDVFIQFYSVDFRGNRVGLA 392

## RESULT 9

A41443  
pepsin (EC 3.4.23.-) precursor, embryonic - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 05-Jun-1992 #sequence\_revision 05-Jun-1992 #text\_change 21-Jul-2000  
C:Accession: A41443  
R:Hayashi, K.; Agata, K.; Mochii, M.; Yasugi, S.; Eguchi, G.; Mizuno, T.  
J. Biochem. 103, 290-296, 1988  
A:Title: Molecular cloning and the nucleotide sequence of cDNA for embryonic chicken pepsin  
A:Reference number: A41443; MUID:88227903  
A:Accession: A41443  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-383 <HAY>  
A:Cross-references: GB:D00215; NID:g2760810; PIDN:BAA00153.1; PID:g222853  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; hydrolase; protein digestion

Query Match 12.3%; Score 298.5; DB 2; Length 383;  
Best Local Similarity 25.2%; Pred. No. 8e-17;  
Matches 90; Conservative 75; Mismatches 125; Indels 67; Gaps 14;

QY 30 YYVEMTVGSPQTLNIIIVDTGSSNFAVGA----APHFLHRYQRQLSSYRDLRKGVYV 85  
Db 76 YGTISITGTPPDFTVVDFTGSSNLWVPVSCTSPACQSHQMNPSQSTYKSTGQNLSI 135  
QY 86 PYTGKWEGLGTLDSLPHGPNVTVRANTAAITSDKFFINGSNWEGILGLAYAEIARP 145  
Db 136 HYTGDMEGTVGCDTVTVASLMDTNQLFGLST-SEPGQFFVY-VKFDGILGLGYPSLAA- 192  
QY 146 DLSLEPPFDSLVRKQTHV-PNLFLSLQLCGAGFPLNQSEVLASVGGSMIIIGDHSLYTGS 204  
Db 193 -DGITPFDNWNESLLEQNLSVYLS-----REPMGSMVYFGGIDSEYFTGSI 240  
QY 205 WYPIRREWYVEIIVRVEINGQDL--KMDCKEYNDKSIIVDSGTTNLRPLKPKVFAAVK 262  
Db 241 NWIPSVQGYQWISDMSIIVNKEIACSSG-----QAIDTGTSLVAGPASDINDIQS 294  
QY 263 SIKAAASSTKFPDGFGLGEOLVCWQAGTTWNIFFPVLSL-----YLMGEVTVNQSPRIILP 318  
Db 295 AVGANQNT-----YGEYSV-----NCSHILAMPVDFVFIIGGI----- 326  
QY 319 QQYLRLPVEDVA---TSQDCYKFAISSSTGTVMGAVIMEGFYVVDRAKRKIGFA 371  
Db 327 -QY--PVPALAYTEONGQGTCTMSSSQNSADLWILGDVFIQFYSVDFRGNRVGLA 380

## RESULT 10

A39314  
gastricisin (EC 3.4.23.3) precursor - bullfrog  
C:Species: Rana catesbeiana (bullfrog)  
C:Date: 19-Jun-1992 #sequence\_revision 19-Jun-1992 #text\_change 22-Jun-1999

C:Accession: A39314  
R:Yakabe, B.; Tanji, M.; Ichinose, M.; Goto, S.; Miki, K.; Kurokawa, K.; Ito, H.; Kag  
J. Biol. Chem. 266, 22436-22443, 1991  
A:Title: Purification, characterization, and amino acid sequences of pepsinogens and  
A:Reference number: A39314; MUID:92042186  
A:Accession: A39314  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-384 <YAK>  
A:Cross-references: GB:M73750; NID:g213687; PIDN:AAA49530.1; PID:g213688  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; hydrolase; protein digestion

Query Match 12.3%; Score 298.5; DB 2; Length 384;  
Best Local Similarity 25.5%; Pred. No. 8.1e-17;  
Matches 100; Conservative 61; Mismatches 142; Indels 89; Gaps 15;

QY 14 SFVEMVDNLKSGGQYVEMTVGSPQTLNIIIVDTGSSNFAVGAAPHPFL-----HR 66  
Db 51 NFATAFEPLANYMDMSYGEISIGTPPQNFVFLVFDGTSSNLWV---PSTYCQSQACTNHP 107  
QY 67 YYQRQLSSTYRDLRKGVYVPTQGWEGELGTLDSLPHGPNVTVRANIA-----AI 118  
Db 108 QFNPSQSSSYSSNQOQFSLQYGTGSLTGILGYDTVQI-----QNIAISQOEFGLSV 158  
QY 119 TESDKFFINGNWEGILGLAYAEIARPDSDLEPFDLSLVKQTHVFN-LFSLQLCGAGFPL 177  
Db 159 TEPGTNFVY-AQFDGILGLAYPSTA--EGGATTVMGIMQNLINQPLFAFLYSGQQNSQ 215  
QY 178 NQSEVLASVGGSMIIIGDHSLYTGSLWYTPIRREWYVEIIVRVEINGQD---LKMDC 234  
Db 216 N-----GGEVAFGGVDQNYYSQIYVTPVSETYQWIGIGQFSVNGQATWCSCQC- 266  
QY 235 EYNVDKSIDVSGTTNLRPLKPKVFAAASIKAAASSTKFPDGFGLGEOLV-CWQAGTTPW 293  
Db 267 ----QGIVDTGTSLTAPQSFVSLMQSIGAQDQN-----GOYAVSCSNITQSLPT 313  
QY 294 NIFPVI-----SLYLMGEVTVNQ---SPRITILPQOYLRPVEDVATSDQDCYKFAISQ 342  
Db 314 ISFTISGVSFPLPPSAVYVQNSGYCIIGIMPTYLPSONGQPL----- 356  
QY 343 SSTGTVMGAVIMEGFYVVDRAKRKIGFAVSA 374  
Db 357 ---WILGDVFLRQYYSVYDLGNNGVGFAPAAA 384

## RESULT 11

C38302  
pepsin (EC 3.4.23.-) II-2/3 precursor - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 14-Jun-1991 #sequence\_revision 14-Jun-1991 #text\_change 23-Feb-1997  
C:Accession: C38302  
R:Kageyama, T.; Tanabe, K.; Koitai, O.  
J. Biol. Chem. 265, 17031-17038, 1990  
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu  
A:Reference number: A38302; MUID:91009127  
A:Accession: C38302  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-387 <KAG>  
A:Cross-references: GB:J05638  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 12.2%; Score 296; DB 2; Length 387;  
Best Local Similarity 26.9%; Pred. No. 1.3e-16;  
Matches 97; Conservative 63; Mismatches 135; Indels 66; Gaps 13;

QY 30 YYVEMTVGSPQTLNIIIVDTGSSNFAVGAAPHPF-----LHRYQRQLSSTYRDLRK 82  
Db 75 YFGTISITGTPPDFTVVDFTGSSNLWV---PSTYCSSLACALHRRFPEDSDSYQGTSET 131



Qy	30	YVYEMTVGSPQPTNLILVDTGSSNFAVGAAPHFF-----LHRYVORQLSSTYRDLRKG	82
		: : :         : :           : :   :       :   :       :	
Db	75	YFGTISGTGPQDFTVFDTGSSNLWY---PSTYCSLACALHKRPENPDSSTYQGTSET	131
Qy	83	VYPYPTQKGWEGELGTDLSIPHGPNVTVRANTAATESDKFF-----INGSNWE	132
		:   :   :   :         : :   :   :   :   :   :   :   :   :	
Db	132	LSITYGTGSMTGILGYDTV-----KVGSIEDTNQIEGLSKTEPGLTFLLFAPED	179
Qy	133	GILGLAAEATARPDDSLPEFFDSLVKQTHV-PNLFSLQLCGAGFPLNQSEVLASVGGSMI	191
		:   :   :   :	
Db	180	GILGLAYPSISSDAT--PVFDNNWNEGLVQDLFSVYLSSDD-----EKGSLVM	227
		:   :   :   :	
Qy	192	IGGIDHSLYTGSLMYTPIRREWYEVIIIVRVEINGQDLKM--DCKEYNDKSIDVSGTTN	249
		:   :   :   :	
Db	228	FGGIDSSYYTGSLNWPVSYEGYQITMDSVSVINGETIACADSC-----QAIVDGTSL	281
		:   :   :   :	
Qy	250	LRLPKKVFEAAVKSIAKASSTKFPDGFWIGEOLV-CWQAGTTPWNIFFPVISLYLMGEVT	308
		:   :   :   :   :	
Db	282	LTGP-----TSAISNIQSYGASK-----NLLGENVISCAIDSLPDIVF-----	321
		:   :   :   :	
Qy	309	NQSFRTIILPQOYLRPVEDVATSQDDCYKFAISQSSTGT-----VMGAVIMEGFFVWF	362
		:   :   :   :	
Db	322	-----TINGIQPLPASAYILKEDDDCTSGLEGMNVDTYTGELWILGDVFIROYTFVD	375
		:   :   :   :	
Qy	363	RARKRIGFAVS	373
		:         :	
Db	376	RANNQLGLAA	386
		:         :	

RESULT 15

S19684

pepsin A (EC 3.4.23.1) 2/3 precursor - Japanese macaque

N;Alternate names: pepsinogen A Isozyme 2/3

C;Species: Macaca fuscata (Japanese macaque)

C;Date: 22-Nov-1993 #sequence\_revision 19-Oct-1995 #text\_change 18-Jun-1999

C;Accession: S19684; S16064

R;Kageyama, T.; Tanabe, K.; Koiwai, O.

Eur. J. Biochem. 202, 205-215, 1991

A;Title: Development-dependent expression of isozymogens of monkey pepsinogen A;Reference number: S19681; PMID: 92037645

F:1-15/Domain: signal sequence #status predicted <SIG>  
F:16-388/Product: pepsinogen A 2/3 #status experimental <PP>  
F:16-388/Domain: activation peptide #status experimental <AP>  
F:363-388/Product: pepsin A 2/3 #status experimental <EN2>  
F:40-41/Cleavage site: Asp-Phe (pepsin) #status experimental  
F:62-63/Cleavage site: Leu-Ile (pepsin) #status experimental  
F:94,277/Active site: Asp #status predicted  
F:97-112,268-272,311-344/Sulfide bonds: #status predicted  
F:107/Binding site: phosphate (Ser) #status predicted  
F:130-131/Domain: catalytic site: phosphatase, phosphoprotein

		Query Match	12.0%	Score 289.5;	DB 1;	Length 388;
		Best Local Similarity	27.0%;	Pred. No. 4.6e-16;		
		Matches 103;	Conservative 66;	Mismatches 141;	Indels 71;	Gaps
QY	3	DEPEPEGRGSGFVEMVDNLRGKSGQGYVYEMVGVSPQTLNLIVDTGSSNFAVGAAPHP	62			
		: : : :		: : : :	: : : :	
Db	64	DEQPLE-----NYLDM-----EYFGTIGTGAQDTFTFDTGSSNLWVPSVCS	108			
		: : : :		: : : :	: : : :	
QY	63	FL-----HRYQRLSSSTYRDLRGKGVYPYTGKWEGLGTDLNYSIPGPNVTVRNTAAI	118			
		: : : :		: : : :	: : : :	
Db	109	SLACTNHNRRNPQDSSTYSQTSSTVSTYTGSGMTGILGYDTVQV---GGISDTNQIFGL	165			
		: : : :		: : : :	: : : :	

QY 119 TESDK-FFINGSNWEGILGLAYAEIARPDDBLEFFEDSLVKQTHV-PNLFSLQLCGAGFP 176  
Db 166 SETEPGSLYYAPFDGILGLAYPSIS--SSGATPVFDNIWNQGLVQDLFSVYLSAD--- 220  
QY 177 LNQSEVLASVGGSMIIGGIDHSLYTGSLWYTPIRREWHYEVIIIVRVEINGODLKMDCKEY 236  
Db 221 -DQS-----GSVIFGGIDSSYTGSLNWPVSVEGYQIISVDSITMNGEATA--CAE- 270  
QY 237 NYDKSIVDSGTTNLRLPKVKVFEAAVKSIIKAASSTKFPDPDGFWLGEQLV-CWQAGTTPWNI 295  
Db 271 -GCQAIVDGTGTSLLTGTPSIANIQSDIGASENSD-----GEMVVSCTAISSLPDIV 321  
QY 296 FPVISLYLMGEVTNQSPRITILPQQYLRPVEDVATSDDDCYK-----FAISQSSTGTVMG 350  
Db 322 F-----TINGIQYVPVPPSAY-----ILQSGSCISGFGMDVPTESGELWILG 364  
QY 351 AVIMEGFYVYVDRARKRIGFA 371  
Db 365 DVFIQYFTVFDNRANNOVGIA 385

Search completed: August 7, 2002, 09:14:17  
Job time: 41 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:13:36 ; Search time 17.62 Seconds  
(without alignments)  
1002.050 Million cell updates/sec

Title: US-09-724-571-43

Perfect score: 2419

Sequence: 1 ETDEPEPGRGSRFVEMVD.....CLRLRQHQDDFADDSILLK 456

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2419	100.0	501	BACE_HUMAN	P56817 homo sapien
2	2392	98.9	501	BACE_RAT	P56819 rattus norv
3	2390	98.8	501	BACE_MOUSE	P56818 mus musculu
4	1171.5	48.4	518	BAE2_HUMAN	Q9y520 homo sapien
5	327	13.5	324	PEP1_GADMO	P56272 gadus morhu
6	314.5	13.0	350	CATD_BOVIN	P80209 bos taurus
7	306	12.6	387	PEP1_RABIT	P28712 oryctolagus
8	302.5	12.5	388	PEP4_MACFU	P27678 macaca fusc
9	302	12.5	367	PEPA_CHICK	P00793 gallus gall
10	299.5	12.4	396	CATE_HUMAN	P14091 homo sapien
11	298.5	12.3	383	PEPE_CHICK	P16476 gallus gall
12	296	12.2	387	PEP2_RABIT	P27821 oryctolagus
13	296	12.2	391	CATE_CAVPO	P25796 cavia porce
14	295.5	12.2	412	CATD_HUMAN	P07339 homo sapien
15	295	12.2	387	PEP4_RABIT	P28713 oryctolagus
16	289.5	12.0	388	PEP2_MACFU	P27677 macaca fusc
17	287	11.9	407	CATD_RAT	P24268 rattus norv
18	286	11.8	387	PEP3_RABIT	P27822 oryctolagus
19	285.5	11.8	398	CATE_RAT	P16228 rattus norv
20	285	11.8	398	CATD_CHICK	Q05744 gallus gall
21	283.5	11.7	388	PEPA_MACMU	P11489 macaca mula
22	282.5	11.7	388	PEPA_HUMAN	P00790 homo sapien
23	281.5	11.6	381	CHYM_SHEEP	P18276 ovis aries
24	281.5	11.6	388	PEP1_MACFU	P03954 macaca fusc
25	279.5	11.6	410	CATD_MOUSE	P18242 mus musculu
26	278.5	11.5	386	PEPA_PIG	P00791 sus scrofa
27	278	11.5	387	PEPA_CALJA	Q9n2d4 callithrix
28	277.5	11.5	396	CATE_RABIT	P43159 oryctolagus
29	277.5	11.5	397	CATE_MOUSE	P10269 mus musculu
30	276.5	11.4	419	CARY_CANAL	P10377 candida alb
31	273.5	11.3	381	CHYM_BOVIN	P00794 bos taurus
32	273	11.3	388	PEP2_RABIT	P27823 oryctolagus
33	270.5	11.2	377	PEP2_MACFU	P03955 macaca fusc

34 270 11.2 376 1 PAG2\_BOVIN Q28057 bos taurus  
35 267.5 11.1 381 1 CHYM\_CALJA Q9n2d2 callithrix  
36 267 11.0 365 1 CATD\_SHEEP Q9mz58 ovis aries  
37 266 11.0 405 1 CARP\_YEAST P07267 saccharomyc  
38 264 10.9 396 1 CARP\_NEUCR Q01294 neurospora  
39 263.5 10.9 388 1 PEPC\_CALJA Q9n243 callithrix  
40 262.5 10.9 394 1 PEPC\_CAVPO Q64411 cavia porce  
41 261.5 10.8 387 1 PEPC\_AEDAE Q03168 aedes aegyp  
42 261.5 10.8 388 1 PEPC\_HUMAN P20142 homo sapien  
43 261 10.8 388 1 PAG\_HORSE Q28389 equus caball  
44 260 10.7 345 1 CATD\_PIG P00795 sus scrofa  
45 255 10.5 496 1 ASPR\_ORYSA P42211 oryza sativ

## ALIGNMENTS

RESULT 1  
BACE\_HUMAN STANDARD; PRT; 501 AA.  
AC P56817; Q9UJT5;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)  
DE (Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl  
DE (protease 2) (Asp 2) (ASP2) (Membrane-associated aspartic protease 2)  
DE (Memapsin-2).  
GN BACE OR BACE1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RC TISSUE=Brain;  
RX MEDLINE=20002972; PubMed=10531052;  
RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,  
RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,  
RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,  
RA Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,  
RA Treanor J., Rogers G., Citron M.;  
RT "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by  
RT the transmembrane aspartic protease BACE.";  
RL Science 286:735-741(1999).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM A), SEQUENCE OF 46-68, AND  
RC CHARACTERIZATION.  
RX TISSUE=Brain;  
RA Sinha S., Anderson J.P., Barbour R., Basl G.S., Caccavello R.,  
RA Davis D., Doan M., Dovey H.F., Frigon N., Hong J., Jacobson-Croak K.,  
RA Jewett N., Keim P., Knops J., Lieberburg I., Power M., Tan H.,  
RA Tatsuno G., Tung J., Schenk D., Seubert P., Suomensaar S.M., Wang S.,  
RA Walker D., Zhao J., McConlogue L., Varghese J.;  
RT "Purification and cloning of amyloid precursor protein beta-secretase  
RT from human brain.";  
RL Nature 402:537-540(1999).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RX MEDLINE=20051710; PubMed=10591213;  
RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,  
RA Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,  
RA Tomasselli A.G., Parodi L.A., Heinrichson R.L., Gurney M.E.;  
RT "Membrane-anchored aspartyl protease with Alzheimer's disease beta-  
RT secretase activity.";  
RL Nature 402:533-537(1999).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RX MEDLINE=20030186; PubMed=10561122;  
RA Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,  
RA Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M.,  
RA Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;

\*Identification of a novel aspartic proteinase (Asp 2) as beta-secretase.\*;  
 RL Mol. Cell. Neurosci. 14:419-427(1999).

[5]  
 RN SEQUENCE FROM N.A. (ISOFORM B).  
 RC TISSUE=Brain, and Pancreas;  
 RA Michel B., De Pietri Tonelli D., Zaccchetti D., Keller P.;  
 RT "New beta-site APP cleaving enzyme isoform (BACE-1B) obtained from  
 human brain and pancreas.";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 [6]

RN SEQUENCE OF 14-501 FROM N.A. (ISOFORM A), AND CHARACTERIZATION.  
 RX MEDLINE=20144060; PubMed=10677483;  
 RA Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;

RT "Human aspartic protease memapsin 2 cleaves the beta-secretase site of  
 beta-amyloid precursor protein.";  
 RL prog. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).

CC -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE  
 CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF  
 CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,  
 CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED  
 CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL  
 CC FRAGMENT WHICH IS LATER RELEASE BY GAMMA-SECRETASE.

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; A/BACE-1A (shown here) and  
 CC B/BACE-1B; are produced by alternative splicing.

CC -!- TISSUE SPECIFICITY: BRAIN.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE

CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL; AF190725; AAF04142.1; -

CC EMBL; AF201468; AAF18982.1; -

CC EMBL; AF200343; AAF17079.1; -

CC EMBL; AF204943; AAF26367.1; -

CC EMBL; AF338816; AAK38374.1; -

CC EMBL; AF200193; AAF13715.1; -

CC HSSP; P32329; LYPS.

CC MIM; 604252; -

CC InterPro: IPR001969; Asp. protease.

CC InterPro: IPR001461; pepsin.

CC Pfam; PF00026; asp. 3.

CC PRINTS; PR00792; PEPSIN.

CC PROSITE; PS00141; ASP\_PROTEASE; 1.

CC KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;

CC Signal; Alternative splicing.

CC FT SIGNAL 1 21 POTENTIAL.

CC FT PROPEP 22 45

CC FT CHAIN 46 501 BETA-SECRETASE.

CC FT DOMAIN 22 457 EXTRACELLULAR (POTENTIAL).

CC FT TRANSMEM 458 478 POTENTIAL.

CC FT DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).

CC FT ACT\_SITE 93 93 BY SIMILARITY.

CC FT ACT\_SITE 289 289 BY SIMILARITY.

CC FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT VARSPLIC 190 214 MISSING (IN ISOFORM B).

CC SEQUENCE 501 AA; 55763 MW; 377CE4C824ACEF05 CRC64;

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

QY 1 ETDEEPEEPRGRGSFVEMVDNLRGKSGQGYVYEMTVGSPQTILNLIYDGTGSSNFAYGAAP 60  
 DB 46 ETDEEPEEPRGRGSFVEMVDNLRGKSGQGYVYEMTVGSPQTILNLIYDGTGSSNFAYGAAP 105  
 QY 61 HPFLHRYQRLSSTYRDLRKGVYVPTQGWEGELGTDLVSPHGPNTVVRANIAAITE 120  
 DB 106 HPFLHRYQRLSSTYRDLRKGVYVPTQGWEGELGTDLVSPHGPNTVVRANIAAITE 165  
 QY 121 SDKFFINGSNWEGILGLAYAEIARPDSDLEPFDFSLVKQTHVNLFSQLCGAGFFLNQS 180  
 DB 166 SDKFFINGSNWEGILGLAYAEIARPDSDLEPFDFSLVKQTHVNLFSQLCGAGFFLNQS 225  
 QY 181 EVLASVGGSMIGIDHSILYTGSLWTPTRREWYEVYIVRVEINGDLKMDCKEYNDK 240  
 DB 226 EVLASVGGSMIGIDHSILYTGSLWTPTRREWYEVYIVRVEINGDLKMDCKEYNDK 285  
 QY 241 SIYDSGTNLRPKVFEAAVKSIAASSTKFPDGFNLGEQLVCVQWAGTTPWNIFFVIS 300  
 DB 286 SIYDSGTNLRPKVFEAAVKSIAASSTKFPDGFNLGEQLVCVQWAGTTPWNIFFVIS 345  
 QY 301 LYLMGEVTNQSFRITILPQOYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYV 360  
 DB 346 LYLMGEVTNQSFRITILPQOYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYV 405  
 QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPPVTLDMEDCGYNIPQTDSTLTMTAYVMAAI 420  
 DB 406 FDRARKRIGFAVSACHVHDEFRTAAVEGPPVTLDMEDCGYNIPQTDSTLTMTAYVMAAI 465  
 QY 421 CALFMPLCLMVCWQRCRLCRLQHQHDDFADDISLLK 456  
 DB 466 CALFMPLCLMVCWQRCRLCRLQHQHDDFADDISLLK 501

RESULT 2

BACKE\_RAT

ID BACKE\_RAT STANDARD; PRT; 501 AA.

AC P56819;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)

DE (Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl

DE protease 2) (Asp 2) (ASP2) (Membrane-associated aspartic protease 2)

DE (Memapsin-2).

GN BACE.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20002972; PubMed=10531052;

RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,

RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,

RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,

RA Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,

RA Treanor J., Rogers G., Citron M.;

RA "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by

the transmembrane aspartic protease BACE.";

RT Science 286:735-741(1999).

RL -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE

CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF

CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,

CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED

CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL

CC FRAGMENT WHICH IS LATER RELEASE BY GAMMA-SECRETASE (BY

CC SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE

CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Query Match 100.0%; Score 2419; DB 1; Length 501;

Best Local Similarity 100.0%; Pred. No. 5.3e-193;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

DR EMBL; AF190727; AAF04144.1; -.
DR HSP; P32329; 1YPS.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF000026; asp.3.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;
FT SIGNAL.
FT PROPEP 1 21 POTENTIAL.
FT CHAIN 22 45 POTENTIAL.
FT DOMAIN 46 501 BETA-SECRETASE.
FT TRANSMEM 22 457 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 458 478 POTENTIAL.
FT DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 93 93 BY SIMILARITY.
FT ACT_SITE 289 289 BY SIMILARITY.
FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 501 AA; 55806 MW; 24B445BC8B87DE3 CRC64;

Query Match 98.9%; Score 2392; DB 1; Length 501;
Best Local Similarity 98.2%; Pred. No. 9.2e-191;
Matches 448; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ETDEPEPEGRGSGFVEMVNLKSGGGYVMTVGGSPQTILNILDVTGSSNFAVGAAP 60
DB 46 ETDEPEPEGRGSGFVEMVNLKSGGGYVMTVGGSPQTILNILDVTGSSNFAVGAAP 105
QY 61 HPFLHRYQRLSTYRDLRKGVVYPTQGWEGELGTDLSVPHGNVTVRANIAITE 120
DB 106 HPFLHRYQRLSTYRDLRKGVVYPTQGWEGELGTDLSVPHGNVTVRANIAITE 165
QY 121 SDRFFINGNWEGILGLAYAEIARPDSDLEPFFDLSLVKQTHVPLFSLQCGAGFPLNQS 180
DB 166 SDRFFINGNWEGILGLAYAEIARPDSDLEPFFDLSLVKQTHVPLFSLQCGAGFPLNQT 225
QY 181 EVLASVGSMILIGIDHSYTGSLWYPIRREWYEVIIIVRVEINGDGLKMDCKEYNDK 240
DB 226 EALASVGSMILIGIDHSYTGSLWYPIRREWYEVIIIVRVEINGDGLKMDCKEYNDK 285
QY 241 SIYDVGTTNLRPKKVFPEAAVKSIAASSTKFPDGFGLGEQLVCWQAGTTPWNIFPVIS 300
DB 286 SIYDVGTTNLRPKKVFPEAAVKSIAASSTKFPDGFGLGEQLVCWQAGTTPWNIFPVIS 345
QY 301 LYLMEVTNOSFRITILPOQLRPVEDVATSDQCYKFAISQSSTGTVMGAVIMEGYVYV 360
DB 346 LYLMEVTNOSFRITILPOQLRPVEDVATSDQCYKFAISQSSTGTVMGAVIMEGYVYV 405
QY 361 FDRARKRIGFAVSACHVHDFRTAAVGGPFVTLDMDCGNIPIQTDSTLMTIAYVMAAI 420
DB 406 FDRARKRIGFAVSACHVHDFRTAAVGGPFVTLDMDCGNIPIQTDSTLMTIAYVMAAI 465
QY 421 CALFMPLCLMVCQWRCRLCRHQHDDFADDISLLK 456
DB 466 CALFMPLCLMVCQWRCRLCRHQHDDFADDISLLK 501

```

RESULT 3

BASE\_MOUSE

ID BASE\_MOUSE

AC P56818;

DT 30-MAY-2000 (Rel. 39, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)
DE (beta-site amyloid precursor protein cleaving enzyme) (Aspartyl
DE protease 2) (Asp 2) (ASP2) (Membrane-associated aspartic protease 2)
DE (Memapsin-2).
DE BACE.
GN BACE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20002972; Pubmed=10531052;
RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,
RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,
RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,
RA Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,
RA Treanor J., Rogers G., Citron M.;
RT "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by
RT the transmembrane aspartic protease BACE.";
RL Science 286:735-741(1999).
RN [2]
RP REVISIONS TO 6 AND 81-87.
RA Bennett B.D., Vassar R., Citron M.;
RA Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057170; Pubmed=10591213;
RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,
RA Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,
RA Tomasselli A.G., Parodi L.A., Heinrikson R.L., Gurney M.E.;
RT "Membrane-anchored aspartyl protease with Alzheimer's disease
RT beta-secretase activity.";
RL Nature 402:533-537(1999).
CC -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE
CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF
CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP.
CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED
CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL
CC FRAGMENT WHICH IS LATER RELEASED BY GAMMA-SECRETASE (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: BRAIN.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

CC EMBL; AF190726; AAF04143.2; -.

CC EMBL; AF200346; AAF17082.1; -.

CC HSP; P56272; 1AM5

CC MGD; MGI:1346542; Bace.

CC InterPro; IPR001969; Asp\_protease.

CC InterPro; IPR001461; Pepsin.

CC Pfam; PF000026; asp. 1.

CC PRINTS; PR00792; PEPSIN.

CC PROSITE; PS00141; ASP\_PROTEASE; 1.

KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;

KW Signal.

FT SIGNAL 1 21 POTENTIAL.

FT PROPEP 22 45 POTENTIAL.

FT CHAIN 46 501 BETA-SECRETASE.

FT DOMAIN 22 457 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 458 478 POTENTIAL.

FT DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).

FT ACT\_SITE 93 93 BY SIMILARITY.

FT ACT\_SITE 289 289 BY SIMILARITY.

FT	CARBOHYD	153	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	172	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	223	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	354	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SQ	SEQUENCE	501 AA; 55747 MW; C085A013145E474E CRC64;		
Query Match 98.8%; Score 2390; DB 1; Length 501;				
Best Local Similarity 98.2%; Pred. No. 1.3e-190;				
Matches 448; Conservative 4; Mismatches 4; Indels 0; Gaps 0;				
QY	1	ETDEEPEEGRGRGSFVEMVDNLKRGKSGQGYVEMTVGSPQTNLNLVDTGSSNFVGAAP	60	
DB	46	ETDEESEPGRGRGSFVEMVDNLKRGKSGQGYVEMTVGSPQTNLNLVDTGSSNFVGAAP	105	
QY	61	HPFLHRYQRLSTYRDLRKGVYVPTQCKWEGELGTDLVSPHGNVTVRANIAAITE	120	
DB	106	HPFLHRYQRLSTYRDLRKGVYVPTQCKWEGELGTDLVSPHGNVTVRANIAAITE	165	
QY	121	SDKFFINGSNWEGILGLAYAEIARPDPSLEPFDDSLVKQTHVPLMSLQLCGAGFPLNQS	180	
DB	166	SDKFFINGSNWEGILGLAYAEIARPDPSLEPFDDSLVKQTHVPLMSLQLCGAGFPLNQT	225	
QY	181	EVLASVGSMIIGIDHSLYTSGIWTYPIRREWYVEIIVRVEINGODLKMCKEYNYDK	240	
DB	226	EALASVGSMIIGIDHSLYTSGIWTYPIRREWYVEIIVRVEINGODLKMCKEYNYDK	285	
QY	241	SIYDSGTTNLRPKKVFEEAAVKSITKAASSTKPDGFWLGEOLVCWQAGTTPNNIPPVIS	300	
DB	286	SIYDSGTTNLRPKKVFEEAAVKSITKAASSTKPDGFWLGEOLVCWQAGTTPNNIPPVIS	345	
QY	301	LYLMGEVTNOSFRITILPOOYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVV	360	
DB	346	LYLMGEVTNOSFRITILPOOYLRPVEDVATSQDDCYKFAVQSSTGTVMGAVIMEGFYVV	405	
QY	361	FDRARKRIGFAVSACHVHDEFTAAVGGPFVTLDMEDCGNIPQTDSTLMTIAYVMAAI	420	
DB	406	FDRARKRIGFAVSACHVHDEFTAAVGGPFVTLDMEDCGNIPQTDSTLMTIAYVMAAI	465	
QY	421	CALFMLPLCLMVCQWRCLRLRQOHHDFADDISLLK	456	
DB	466	CALFMLPLCLMVCQWRCLRLRQOHHDFADDISLLK	501	
RESULT 4				
BAE2_HUMAN				
ID	BAE2_HUMAN	STANDARD;	PRT;	518 AA.
AC	Q9Y520; Q9UT6;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Beta secretase 2 precursor (EC 3.4.23.-) (Beta-site APP-cleaving			
DE	enzyme 2) (Aspartyl protease 1) (Asp 1) (ASPL) (Membrane-associated			
DE	aspartic protease 1) (Memapsin-1).			
GN	BACE2 OR ASP21.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20057170; PubMed=10591213;			
RA	Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,			
RA	Brashler J.R., Stratman N.C., Matthews W.R., Buhl A.E., Carter D.B.,			
RA	Tomasselli A.G., Parodi L.A., Heinrikson R.L., Gurney M.E.;			
RT	"Membrane-anchored aspartyl protease with Alzheimer's disease			
RT	beta-secretase activity."			
RL	Nature 402:533-537(1999).			
RP	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUP=Bone marrow;			
RA	Xin H., Stephens J.C., Duan X., Harrowe G., Kim E., Grieshammer U.,			
RA	Gliese K.;			

RT	"Identification of a novel aspartic-like protease differentially			
RT	expressed in human breast cancer cell lines."			
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.			
RP	SEQUENCE FROM N.A.			
RA	Accarino M.P., Fumagalli P., Ottolenghi S., Taramelli R.;			
RT	"Cloning of a gene from chromosome 21 down region encoding a potential			
RT	transmembrane aspartyl protease."			
RL	Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.			
RP	SEQUENCE FROM N.A.			
RA	Solans A., Estivill X., de la Luna S.;			
RT	"Cloning of a novel mammalian aspartyl protease."			
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.			
RP	SEQUENCE FROM N.A.			
RA	Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,			
RA	Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M.,			
RA	Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;			
RT	"Identification of a novel aspartic proteinase (Asp 2) as			
RT	beta-secretase."			
RL	Mol. Cell. Neurosci. 14:419-427(1999).			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=20144060; PubMed=10677483;			
RA	Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;			
RT	"Human aspartic protease memapsin 2 cleaves the beta-secretase site of			
RT	beta-amyloid precursor protein."			
RL	Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).			
RP	SEQUENCE FROM N.A.			
RA	Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,			
RA	Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,			
RA	Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,			
RA	Manzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,			
RA	Reichwald A., Kump J., Shibuya K., Kawasaki K., Asakawa S.,			
RA	Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,			
RA	Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,			
RA	Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,			
RA	Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,			
RA	Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,			
RA	Lehrach H., Reinhardt R., Yaspo M.-L.;			
RT	"The DNA sequence of human chromosome 21."			
RL	Nature 405:311-319(2000).			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE			
CC	EUKARYOTIC ASPARTYL PROTEASES FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
CC	EMBL; AF200342; AAF17078.1; -			
DR	EMBL; AF117892; AAD45240.1; -			
DR	EMBL; AF050171; AAD45963.1; -			
DR	EMBL; AF178532; AAF29494.1; -			
DR	EMBL; AF204944; AAF26368.1; -			
DR	EMBL; AF200192; AAF13714.1; -			
DR	EMBL; AL163284; CAB90458.1; -			
DR	EMBL; AL163285; CAB90554.1; -			
DR	HSSP; P00797; 2REN.			
DR	MM; 605668; -			
DR	InterPro; IPR001969; Asp_protease.			
DR	InterPro; IPR001461; Pepsin.			
DR	Pfam; PF00026; asp; 3.			
DR	PRINTS; PR00792; PEPSIN.			



```
RL FEBS Lett. 319:54-58(1993).
RN [2]
RP SEQUENCE OF 45-390, AND X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
RC TISSUE=Liver;
RA MEDLINE=93223670; PubMed=8467789;
RX Metcalf P., Fusek M.;
RT "Two crystal structures for cathepsin D: the lysosomal targeting
RL signal and active site.";
RL EMBO J. 12:1293-1302(1993).
CC -1- FUNCTION: CATHEPSIN D IS AN ACID PROTEASE ACTIVE IN INTRACELLULAR
CC PROTEIN BREAKDOWN.
CC -1- CATALYTIC ACTIVITY: Specificity similar to, but narrower than,
CC that of pepsin A. Does not cleave the 4-Gln-His-5 bond in B
CC chain of insulin.
CC -1- SUBUNIT: CONSISTS OF A LIGHT CHAIN AND A HEAVY CHAIN.
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PIR; S32383; S32383.
DR PIR; S37419; S37419.
DR HSP; P07339; 1LYB.
DR MEROPS; A01.009; -.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PS00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Glycoprotein; Lysosome; Zymogen.
FT PROPEP 1 44
FT CHAIN 45 390
FT ACT_SITE 77 77
FT ACT_SITE 273 273
FT DISULFID 71 140
FT DISULFID 90 97
FT DISULFID 264 268
FT DISULFID 307 344
FT CARBOHYD 114 114
FT CARBOHYD 241 241
FT SEQUENCE 390 AA; 42488 MW; 5B38AA1C33C48D35 CRC64;
Query Match 13.0%; Score 314.5; DB 1; Length 390;
Best Local Similarity 28.0%; Pred. No. 1.3e-18;
Matches 113; Conservative 72; Mismatches 128; Indels 91; Gaps 21;
QY 8 EPG--RRGSFVEMVDNLGRKSGQGYVEMTVGSPQTNLILVDTGSSNFAYGAAPHPFL-- 64
DB 39 EPAVRQGPPELLKNYDAQ---YYGEIGIGTPQCFTVVFDGTSANLWVPSIHCKLLDI 95
QY 65 ----HRYQRLSTYRDLRKGVY--VPYTGKKEGELGTDLVSIPIGPN-----VTVR 112
DB 96 ACWTHRYKNSDKSSTY--VKNGTTFDIHYGSGSLGYSQDTSVPCNPSSSPGGVTVQ 153
QY 113 ANI--AAITESDKFINGSNWEGILGLAYAEIARPDLSLEPFDSLVAIKOTHV--PNLFSIQ 169
DB 154 RQTFGEAIKQGVVFI--AAKFDGILGMAYPRIS--VNNVLGFDNLQCKLVDRKNFS-- 208
QY 170 LCGAGFPLNQSLEVASVGGSMIIGCIDHSILYTGSLWTPTIRREWYEVIIVRVEINGQDL 229
DB 209 -----FFLNLR-DPKAQPGGELMLGTDTSKYRGSLMFHNTVROAYWQIHMQLDV-GSSL 261
QY 230 KMDCKEYNYKSIYDSTGNTNLRPKVFEAAVKAISAKSTEKFPDGFWEQLV-CWQA 288
DB 262 TV-CK--GGCEAIYDVTGSLTVGPVEEVREIQAIGAVPLIQ-----GEYMIPCEK 310
QY 289 GTTPNFPVLSILMGEVNTQSFRIITLPOQLRPVEDVATSDCCYKFAISOSSTGT- 347
DB 311 SS-----LPEVTVKLGG-----KDYALSPED-YALKVSAQETVC 344
QY 348 -----VMGAVIMEGFVYVDFDRARKRIGFAVSA 374
DB 345 LSGFMGMDIPPPGGLWILGDFVIGRYVTVFDRQNRVGLAEAA 388
```

```
RESULT 7
PEP1_RABIT STANDARD; PRT; 387 AA.
ID PEP1_RABIT
AC P28712;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Pepsin II-1 precursor (EC 3.4.23.1) (Pepsin A).
OS Oryctolagus cuniculus (rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91009127; PubMed=2129536;
RA Kageyama T., Tanabe K., Koizumi O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. Biol. Chem. 265:17031-17038(1990).
CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -1- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PIR; B38302; B38302.
DR HSP; P00791; 1PSA.
DR MEROPS; A01.001; -.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
DR Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW Zymogen; Signal; Phosphorylation; Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 59
FT CHAIN 60 387
FT MOD_RES 129 129
FT ACT_SITE 93 93
FT ACT_SITE 276 276
FT DISULFID 106 111
FT DISULFID 267 271
FT DISULFID 310 343
FT SEQUENCE 387 AA; 42070 MW; A6EC48F715541A48 CRC64;
Query Match 12.6%; Score 306; DB 1; Length 387;
Best Local Similarity 27.1%; Pred. No. 6.4e-18;
Matches 98; Conservative 67; Mismatches 131; Indels 66; Gaps 15;
QY 30 YYVEMTVGSPQTNLILVDTGSSNFAYG----AAPHPLHRYQRLSTYRDLRKGVYV 85
DB 75 YFGTISGTTPQETVIFDGTGSSNLWVPSTYCCSLACFLHKRFPDSDSTFQATSETLSI 134
QY 86 PYTGKKEGELGTDLVSIPIGPNVTVRANITAEISD---KFFINGSNWEGILGLAYAEI 142
DB 135 TYGTGSMTGILGYDTVKV---GNIEDTNQIFGLSKTEPGITFLV--APFDGILGLATPSI 189
QY 143 ARPDLSLEPFDSLVAIKOTHV--PNLFSLQLCAGFPLNQSEVLASVGGSMIIGCIDHSILY 201
DB 190 SASDAT--PFDNNWNEGLYSEDLFSVYLSNG-----EKGSMVMEGIDSSVYT 237
QY 202 GSLWYTPFIRREWYEVIIVRVEINGQDLKM--DCKEYNYKSIYDSTGNTNLRPKVFEA 259
DB 238 GSNLWVPVSEGYWQITMDSITINGETIACADSC-----QAVYDTGTSLLAGPTSAISK 291
```









```

CC      EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; M59235; AAA85369.1; -.
CC      PIR; C38302; C38302.
CC      HSP; P00790; 1PSN.
CC      MEROPS; A01.001; -.
CC      InterPro; IPR001969; Asp_protease.
CC      Pfam; PF00026; asp; 1.
CC      PRINTS; PR00792; PEPsin.
CC      PROSITE; PS00141; ASP_PROTEASE; 2.
CC      HydroLase; Aspartyl protease; Digestion; Stomach; Gastric juice;
CC      Zymogen; Signal; Phosphorylation; Multigene family.
CC      SIGNAL 1 15
CC      PROPEP 16 59 ACTIVATION PEPTIDE.
CC      CHAIN 60 387 PEPsin II-2/3.
CC      MOD_RES 129 129 PHOSPHORYLATION (POTENTIAL).
CC      ACT_SITE 93 93 BY SIMILARITY.
CC      ACT_SITE 276 276 BY SIMILARITY.
CC      DISULFID 106 111 BY SIMILARITY.
CC      DISULFID 267 271 BY SIMILARITY.
CC      DISULFID 310 343 BY SIMILARITY.
CC      SEQUENCE 387 AA; 42100 MW; 66FC331A3DC75891 CRC64;

Query Match 12.2%; Score 296; DB 1; Length 387;
Best Local Similarity 26.9%; Pred. No. 4.4e-17;
Matches 97; Conservative 63; Mismatches 135; Indels 66; Gaps 13;

Qy 30 YVEMTVGSPPTLNILVDVTGSSNFVAGAAPHP-----LHRYQRLSSSTYRDLRKG 82
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 75 YFGTISTGTPQDFTVIFDTGSSNLWV---PSTYCSSLACALHKRFNPEDSSIVQGTSET 131
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 83 YVYPYTGKWEGLGTDLVSIHPGPNVTVRANIAITAIESDKFFINGSNWEGILGLAYAEI 142
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 132 LSIYTGTSMTGILGYTVKVGSTIEDNQIFGLSKTFPSLTFLF--APFDGILGLAYPSI 189
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 143 ARPDLSLEPFDSLAVKTHV--PNLFSLQLCAGAPLQSEVLASVSGSMIIGGDHSLYT 201
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 190 SSDAT--PVFDNNWNEGLVSDQLFSVILSSDD-----EKGSLVFEGLDSSYYT 237
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 202 GSLWYTPIRREWEYEVIVRVEINGODLKM--DCKEYNKDSIVDSGTTNLRLPKKVFEA 259
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 GSLNWPVSVYEGYQWITMDSVISINGETIACADSC-----QAIVDTGTSLLTGP---TS 287
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 260 AVSIIKAASSTKPEPDGFWGLGQVY-CWQAGTTPNWIFPVVLSLIMGEVTVNQSFRIILP 318
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 288 AISNIQSYIGASK-----NLIGENVISCAIDSLDPIVF-----TING 325
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 319 QOYLRPVEDVATSDCCYKFAISOSSTGT-----VMGAVIMEGFYVVFVRKRKRIGFAV 372
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 326 IQYPLPASAYILKEDDCTSCGLEGMNVDTYTGELWILGDFVIQYFYVFDNRANQLGLAA 385
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 373 S 373
Db 386 A 386

RESULT 13
CATE_CAVPO STANDARD; PRT; 391 AA.
ID CATE_CAVPO
AC P25796;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Cathepsin E precursor (EC 3.4.23.34).
CTSE.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Mystricognathi; Cavidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=92355614; PubMed=1644829;
RA Kageyama T., Ichinose M., Tsukada S., Miki K., Kurokawa K., Koizumi O.,
RA Tanji M., Yakabe E., Athuda S.B., Takahashi K.;
RT "Gastric procathepsin E and progastricatin from guinea pig.
RT Purification, molecular cloning of cDNAs, and characterization of
RT enzymatic properties, with special reference to procathepsin E.";
RL J. Biol. Chem. 267:16450-16459(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Gastric mucosa;
RP MEDLINE=96073637; PubMed=8540321;
RA Kageyama T., Ichinose M., Miki K., Moriyama A., Yonezawa S.,
RA Tanji M., Athuda S.B., Takahashi K.;
RT "Isolation, characterization, and structure of procathepsin E and
RT cathepsin E from the gastric mucosa of guinea pig.";
RL Adv. Exp. Med. Biol. 362:211-221(1995).
CC -!- FUNCTION: DUE TO ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN
CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.
CC -!- CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader
CC specificity.
CC -!- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M88653; AAA37052.1; -.
CC EMBL; S80547; AAB35844.1; -.
CC PIR; A43356; A43356.
CC HSP; P00794; 4CMS.
CC MEROPS; A01.010; -.
CC InterPro; IPR001969; Asp_protease.
CC Pfam; PF00026; asp; 1.
CC PRINTS; PR00792; PEPsin.
CC PROSITE; PS00141; ASP_PROTEASE; 2.
CC HydroLase; Aspartyl protease; Glycoprotein; Zymogen; Signal.
CC SIGNAL 1 18 POTENTIAL.
CC PROPEP 19 53 ACTIVATION PEPTIDE (POTENTIAL).
CC CHAIN 54 391 CATHEPSIN E.
CC ACT_SITE 92 92 BY SIMILARITY.
CC ACT_SITE 276 276 BY SIMILARITY.
CC DISULFID 56 56 INTERCHAIN (PROBABLE).
CC DISULFID 105 110 BY SIMILARITY.
CC DISULFID 267 271 BY SIMILARITY.
CC DISULFID 309 346 BY SIMILARITY.
CC CARBOHYD 86 86 N-LINKED (GLCNAC... ) (POTENTIAL).
CC SEQUENCE 391 AA; 42132 MW; 78D216BF8CFDABD CRC64;

Query Match 12.2%; Score 296; DB 1; Length 391;
Best Local Similarity 26.9%; Pred. No. 4.4e-17;
Matches 98; Conservative 64; Mismatches 130; Indels 72; Gaps 16;

Qy 30 YVEMTVGSPPTLNILVDVTGSSNFVAG---APHFLHRYQRLSSSTYRDLRKGYYV 85
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 74 YFGTISTGSPQNFVTIVFDTGSSNLWVPSVYCTSPACQTHPVFHPSPSSSTYRENGNSFSI 133
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 86 PYTGKWEGLGTDLVSIHPGPNVTVRANIAITAIESDKFFINGSNWEGILGLAYAEIARP 145

```

```
Db 134 QYGTSLTGIIGADQSV-EGLTVVGOQFGEVQEPGKTEVH-AEFDGLILGLYPSLAA- 190
Qy 146 DDLSEPFDSLQKQHPVNLPSLQCLGAGFPLNQSEVLASVGGSMIIGDHSLYTGSIM 205
Db 191 -GGVTPVDFNMMAQ-----NLVALPM-----FSYMSNPGSGSELTFGGYDPFSHFGSLN 241
Qy 206 YTPIRREWYEVIIIVRVEINGODLKMDCKEYNDKSIYDSGTNTLRLPKVFEAAVKSIK 265
Db 242 WPVTKQAYWQIALDGIQV--DSVMFCSE--GCQAIYDTGTSITGP-----PGKIKQLQ 293
Qy 266 AASSTEKPFDFGLGEQLVQWQATTPWNIPIVISLYLMGEVYNOSFRI-----TILPQQ 320
Db 294 EALGATYVDEGY-----SVQC-----ANLNMLDVT-----FIINGVPYTLNPTA 333
Qy 321 YLRPVEDVATSDCCYKFAISQSSFG-----TWGAVIMEGFVVVEDRKR 367
Db 334 Y--TLLDFVDMQVC-----STGFEGLEIQPPAGPLWILGDVFIHQFVAFVDRGNR 383
Qy 368 IGFA 371
Db 384 VGLA 387

RESULT 14
CATD_HUMAN
ID CATD_HUMAN STANDARD; PRT; 412 AA.
AC P07339;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cathepsin D precursor (EC 3.4.23.5).
OS CTSD.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85270436; PubMed=3927292;
RA Faust P.L., Kornfeld S., Chirgwin J.M.;
RT "Cloning and sequence analysis of cDNA for human cathepsin D.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4910-4914(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87231068; PubMed=3588310;
RA Westley B.R., May F.E.B.;
RT "Oestrogen regulates cathepsin D mRNA levels in oestrogen responsive
human breast cancer cells.";
RL Nucleic Acids Res. 15:3773-3786(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91299158; PubMed=2069717;
RA Redeker B., Heckendorf B., Grosch H.W., Mersmann G., Hasilik A.;
RT "Molecular organization of the human cathepsin D gene.";
RL DNA Cell Biol. 10:423-431(1991).
RN [4]
RP SEQUENCE OF 1-22 FROM N.A.
RX MEDLINE=94085791; PubMed=8262386;
RA May F.E., Smith D.J., Westley B.R.;
RT "The human cathepsin D-encoding gene is transcribed from an estrogen-
regulated and a constitutive start point.";
RL Gene 134:277-282(1993).
RN [5]
RP SEQUENCE OF 1-22 FROM N.A.
RX MEDLINE=95021301; PubMed=7935485;
RA Augereau P., Miralles F., Cavailles V., Gaudelet C., Parker M.,
Rocheport H.;
RT "Characterization of the proximal estrogen-responsive element of
human cathepsin D gene.";
RL Mol. Endocrinol. 8:693-703(1994).
RN [6]
RP SEQUENCE OF 170-180.
```

```
RC TISSUE=Liver;
RA Hochstrasser D.F., Frutiger S., Paquet N., Bairoch A., Ravier F.,
RA Pasquelli C., Sanchez J.-C., Tissot J.-D., Bjellqvist B., Vargas R.,
RA Appel R.D., Hughes G.J.;
RL Submitted (JUN-1992) to the SWISS-PROT data bank.
RN [7]
RP VARIANT VAL-58.
RX MEDLINE=20179010; PubMed=10716266;
RA Papasotiropoulos A., Bagli M., Kurz A., Kornhuber J., Forstl H.,
RA Maier W., Pauls J., Lautenschlager N., Heun R.;
RT "A genetic variation of cathepsin D is a major risk factor for
Alzheimer's disease.";
RL Ann. Neurol. 47:399-403(2000).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
RC TISSUE=Spleen;
RX MEDLINE=93223670; PubMed=8467789;
RA Metcalf P., Fusek M.;
RT "Two crystal structures for cathepsin D: the lysosomal targeting
signal and active site.";
RL EMBO J. 12:1293-1302(1993).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RC TISSUE=Liver;
RX MEDLINE=93342076; PubMed=8393577;
RA Baldwin E.T., Bhat T.N., Gulnik S., Hosur M.V., Sowder R.C. II,
RA Cachau R.E., Collins J., Silva A.M., Erickson J.W.;
RT "Crystal structures of native and inhibited forms of human cathepsin
D: implications for lysosomal targeting and drug design.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:6796-6800(1993).
CC -!- FUNCTION: CATHEPSIN D IS AN ACID PROTEASE ACTIVE IN INTRACELLULAR
PROTEIN BREAKDOWN.
CC -!- CATALYTIC ACTIVITY: Specificity similar to, but narrower than,
that of pepsin A. Does not cleave the 4-Gln-1-His-5 bond in B
chain of insulin.
CC -!- SUBUNIT: CONSISTS OF A LIGHT CHAIN AND A HEAVY CHAIN.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- POLYMORPHISM: The Val-58 allele is significantly overrepresented
in demented patients (11.8%) compared with nondemented controls
(4.9%). Carriers of the Val-58 allele have a 3.1-fold increased
risk for developing AD than noncarriers.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
EUKARYOTIC ASPARTYL PROTEASES FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M11233; AAB59529.1; -
EMBL; X05344; CAA28955.1; -
EMBL; M63138; AAA51922.1; -
EMBL; M63134; AAA51922.1; JOINED.
EMBL; M63135; AAA51922.1; JOINED.
EMBL; M63136; AAA51922.1; JOINED.
EMBL; M63137; AAA51922.1; JOINED.
EMBL; L12980; AAA16314.1; -
EMBL; S74689; AAD14156.1; -
EMBL; S52557; AAD13868.1; -
PIR; A25771; KKHUO.
PDB; 1LYA; 31-JAN-94.
PDB; 1LYB; 31-JAN-94.
MEROPS; A01.009; -
SWISS-2DPAGE; P07339; HUMAN.
Sienna-2DPAGE; P07339; -
MIM; 116840; -
InterPro; IPR001969; Asp_protease.
InterPro; IPR001461; pepsin.
Pfam; PF00026; asp; 1.
PRINTS; PR00792; PEPSIN.
```



```
DR MEROPS; A01.001; -
DR InterPro; IPR001969; Asp.protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp_1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW Zymogen; Signal; Phosphorylation; Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 59 ACTIVATION PEPTIDE.
FT CHAIN 60 387 PERSIN II-4.
FT MOD_RES 129 129 PHOSPHORYLATION (POTENTIAL).
FT ACT_SITE 93 93 BY SIMILARITY.
FT ACT_SITE 276 276 BY SIMILARITY.
FT DISULFID 106 111 BY SIMILARITY.
FT DISULFID 267 271 BY SIMILARITY.
FT DISULFID 310 343 BY SIMILARITY.
SQ SEQUENCE 387 AA; 42052 MW; 21ADD07782A89585 CRC64;

Query Match 12.2%; Score 295; DB 1; Length 387;
Best Local Similarity 26.1%; Pred.No. 5.2e-17;
Matches 97; Conservative 65; Mismatches 123; Indels 86; Gaps 14;

QY 30 YVEMTVGSPQTNLILVDGTSSNFAYGAAPHF-----LHRYQROLSSTYRDLRKG 82
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
75 YFGTISIGTPQDFTVIFDTGSSNLWV---PSTYCSSLACALHKRNPEDSSTYQGTSET 131
QY 83 VVYPYTGKWEGLGTDLVSPHCPNVTVRANIAAITESDKFF-----INGSNWE 132
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
132 LSTYGTGSMTGILGYDTV-----KVGSIEDTNQIFGLSKTEPGLTFLFAPFD 179
QY 133 GILGLAYAEIARPDSDLPPFDLSLVKOTHV-PNLFSLQLCGAGFPLNQSEVLASVGGSMI 191
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
180 GILGLAYPSISSDAT--PVEDNNWNEGLVSQDLFSVLSDD-----EKGLVM 227
QY 192 IGGIDHSLYTGSLWYTPIRREWEYEVIVRVEINGQDLKM--DCKEYNDKSIYVDSGTIN 249
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
228 FGGIDSSYVTGSLNWPVSYEGYQWITMDSVSVINGETIACADSC-----QAIVDTCGTS 281
QY 250 LRLPKKYFEAAVKAASIKASTEKEFPDGFWLGEQLV-CWQAGTTWNIFPVISLYLMGEVT 308
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
282 LTGP-----TSAISNIQSYIGASK-----NLGENVISCASIDSLPDIVF----- 321
QY 309 NQSPRIITLPQQYLRPVEDVATSDDCYKFAISQSSTGT-----VMGAVIMEGFYVVED 362
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
322 -----TINGIQPLPASAYILKEDDDCTSGLEGMNVDTYTGELWILQDWFIRQYFTVFD 375
QY 363 RARRIGFAVS 373
Db : : : : :
376 RANNQLGLAAA 386
```

Search completed: August 7, 2002, 09:15:11  
Job time: 95 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:13:36 ; Search time 50.29 Seconds  
(without alignments)  
1568.616 Million cell updates/sec

Title: US-09-724-571-43  
Perfect score: 2419  
Sequence: 1 ETDEPEPEGRGSGFVEMVD.....CLRLRQHQHDFADDISLLK 456

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rhodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match %	Length DB ID	Description	
1	2419	100.0	532 4	Q9ULS1	Q9ULS1 homo sapien
2	2264.5	93.6	476 4	Q9BYC1	Q9BYC1 homo sapien
3	2161	89.3	457 4	Q9BYC0	Q9BYC0 homo sapien
4	2016.5	83.4	432 4	Q9BYB9	Q9BYB9 homo sapien
5	1407	58.2	266 11	Q9C0U5	Q9C0U5 mus musculus
6	1160.5	48.0	439 4	Q9H2V8	Q9H2V8 homo sapien
7	1146.5	47.4	514 11	Q9JL18	Q9JL18 mus musculus
8	972.5	40.2	468 4	Q9NZL2	Q9NZL2 homo sapien
9	967.5	40.0	396 4	Q9NZL1	Q9NZL1 homo sapien
10	712.5	29.5	213 4	Q9P0D2	Q9P0D2 homo sapien
11	596.5	24.7	255 11	Q9R1P7	Q9R1P7 mus musculus
12	332.5	13.7	391 5	Q9VKP6	Q9VKP6 drosophila
13	332	13.7	354 5	Q9GYX7	Q9GYX7 boophilus m
14	312.5	12.9	386 6	Q9BGU5	Q9BGU5 bos taurus
15	308	12.7	384 13	Q9DEC2	Q9DEC2 xenopus lae
16	308	12.7	385 13	Q9DEC4	Q9DEC4 rana catesb

17	305	12.6	386	6	Q9GMV7	Q9gm7 rhinolophus
18	305	12.6	387	6	Q9GMV8	Q9gm8 sorex ungui
19	304.5	12.6	372	5	Q9VLK3	Q9vlk3 drosophila
20	304.5	12.6	387	13	Q9DDV5	Q9ddv5 salvelinus
21	304	12.6	387	6	Q9GMV9	Q9gm9 suncus murl
22	303.5	12.5	383	13	Q9DE45	Q9de45 salvelinus
23	303	12.5	383	13	Q9DEC3	Q9dec3 xenopus lae
24	302.5	12.5	376	13	Q9PUR8	Q9pur8 pseudopleur
25	301	12.4	382	13	Q9PRG9	Q9prg9 gallus gall
26	301	12.4	423	5	Q9VKP7	Q9vkp7 drosophila
27	298.5	12.3	384	13	Q9I322	Q9i322 rana catesb
28	295.5	12.2	386	6	Q9GMV6	Q9gm6 canis famil
29	294	12.2	396	13	Q93428	Q93428 chionodraco
30	290.5	12.0	381	6	Q9GK11	Q9gk11 camelus dro
31	288	11.9	399	13	Q93458	Q93458 podarcis si
32	287.5	11.9	444	5	Q21966	Q21966 caenorhabdi
33	284	11.7	398	13	P87370	P87370 oncorhynch
34	284	11.7	427	5	P91802	P91802 schistosoma
35	281	11.6	378	13	Q9PUR9	Q9pur9 pseudopleur
36	279.5	11.6	390	6	Q9GK10	Q9gk10 camelus dro
37	278	11.5	370	6	Q9TTW1	Q9ttw1 bos taurus
38	278	11.5	399	13	Q9DD89	Q9dd89 brachydanio
39	278	11.5	422	5	Q96906	Q96906 onchocerca
40	277.5	11.5	380	6	Q28950	Q28950 sus scrofa
41	277	11.5	446	5	Q9N9H3	Q9n9h3 necator ane
42	276	11.4	389	13	Q9PWK1	Q9pwk1 gallus gall
43	276	11.4	389	13	Q9W643	Q9w643 gallus gall
44	276	11.4	396	13	Q9DEX3	Q9dex3 clipea hare
45	275	11.4	385	6	Q29080	Q29080 sus scrofa

## ALIGNMENTS

RESULT 1

Q9ULS1 PRELIMINARY; PRT; 532 AA.

ID Q9ULS1

AC Q9ULS1; 01-MAY-2000 (TRENBLrel. 13, Created)

DT 01-MAY-2000 (TRENBLrel. 18, Last sequence update)

DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)

DE KIAA1149 PROTEIN (FRAGMENT).

GN KIAA1149.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BRAIN;

RX MEDLINE=20039618; PubMed=10574461;

RA Hirotsawa M., Nagase T., Ishikawa K., Kikuno R., Ohara O.;

RT "Characterization of cDNA clones selected by the GeneMark analysis

RL from size-fractionated cDNA libraries from human brain.";

DR EMBL; AB032975; BAA86463.2; --

DR HSSP; P56272; IAM5.

DR MEROPS; A01.004; --

DR InterPro; IPR001969; Asp\_protease.

DR InterPro; IPR001461; Pepsin.

DR Pfam; PF00026; asp; 1.

DR PRINTS; PS00792; PEPsin.

DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.

FT NON\_TER 1

SQ SEQUENCE 532 AA; 58720 MW; 98B135D0D5FBD2E8 CRC64;

Query Match 100.0%; Score 2419; DB 4; Length 532;  
Best Local Similarity 100.0%; Pred. No. 7.3e-198;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ETDEPEPEGRGSGFVEMVDNLKSGQGYVEMTVGSPPTNLIVDTGSSNFVGAAP 60  
|||||

106 HPFLHRYQRLSSTYRDLRKGVVVPTQCKWEGELGTLVSIPIHGPNTVTRANIAAITE 165  
121 SDKFFINGNWEGILGLAYAEIARPDLSLPPFDLSLVKQTHVPNLFSLQCGAGFPLNQS 180  
166 SDKFFINGNWEGILGLAYAEIAR-----LCGAGFPLNQS 200  
181 EVLASVGSMTIIGDHSLSLTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDK 240  
201 EVLASVGSMTIIGDHSLSLTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDK 260  
241 SIVDSGTTNLRPKKVFEEAAVKSIAKASSTKFPDGFGLGEOLVCWQAGTTPWNIFFVIS 300  
261 SIVDSGTTNLRPKKVFEEAAVKSIAKASSTKFPDGFGLGEOLVCWQAGTTPWNIFFVIS 320  
301 LYLAGEVNTQSFRTILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVV 360  
321 LYLAGEVNTQSFRTILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVV 380  
361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDSTLMTIAYVMAAI 420  
381 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDSTLMTIAYVMAAI 440  
421 CALFMLPLCLMWQWRCRLRCLRQHQHDDFADDISLLK 456  
441 CALFMLPLCLMWQWRCRLRCLRQHQHDDFADDISLLK 476

RESULT 3  
Q9BYC0 PRELIMINARY; PRT; 457 AA.  
AC Q9BYC0; PRELIMINARY; PRT; 457 AA.  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE BETA-SITE APP CLEAVING ENZYME I-457 (BETA-SITE APP CLEAVING ENZYME TYPE C).  
DE BACE.  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRIN;  
RX MEDLINE=21408467; PubMed=11516562;  
RA Tanahashi H., Tabira T.;  
RT "Three novel alternatively spliced isoforms of the human beta-site APP cleaving enzyme (BACE) and their effect on amyloid beta-peptide production.";  
RL Neurosci. Lett. 307:9-12(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EXOCRINE PANCREAS;  
RX Zaccchetti D., De Pietri Tonelli D., Schnurbus R.;  
RT "New beta-site APP cleaving enzyme isoform (BACE-1C) obtained from human pancreas.";  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB050437; BAB40932.1; -.  
DR EMBL; AF338817; AAK38375.1; -.  
DR HSSP; P32329; 1YPS.  
DR InterPro; IPR001969; Asp\_protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF00026; asp; 1.  
DR PRINTS; PR00792; PEPsin.  
DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
SQ SEQUENCE 457 AA; 51068 MW; C794C9A9E85FE7A2 CRC64;

Query Match 89.3%; Score 2161; DB 4; Length 457;  
Best Local Similarity 90.4%; Pred. No. 6.1e-176;  
Matches 412; Conservative 0; Mismatches 0; Indels 44; Gaps 1;  
QY 1 ETDEEPEEPCRRGSGFVEMVDNLRKSGQYVYVEMTVGSPQTLNILDVTGSSNFAVGAAP 60

77 ETDEEPEEPCRRGSGFVEMVDNLRKSGQYVYVEMTVGSPQTLNILDVTGSSNFAVGAAP 136  
61 HPFLHRYQRLSSTYRDLRKGVVVPTQCKWEGELGTLVSIPIHGPNTVTRANIAAITE 120  
137 HPFLHRYQRLSSTYRDLRKGVVVPTQCKWEGELGTLVSIPIHGPNTVTRANIAAITE 196  
121 SDKFFINGNWEGILGLAYAEIARPDLSLPPFDLSLVKQTHVPNLFSLQCGAGFPLNQS 180  
197 SDKFFINGNWEGILGLAYAEIARPDLSLPPFDLSLVKQTHVPNLFSLQCGAGFPLNQS 256  
181 EVLASVGSMTIIGDHSLSLTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDK 240  
257 EVLASVGSMTIIGDHSLSLTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDK 316  
241 SIVDSGTTNLRPKKVFEEAAVKSIAKASSTKFPDGFGLGEOLVCWQAGTTPWNIFFVIS 300  
317 SIVDSGTTNLRPKKVFEEAAVKSIAKASSTKFPDGFGLGEOLVCWQAGTTPWNIFFVIS 376  
301 LYLAGEVNTQSFRTILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVV 360  
377 LYLAGEVNTQSFRTILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVV 436  
361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDSTLMTIAYVMAAI 420  
437 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDSTLMTIAYVMAAI 496  
421 CALFMLPLCLMWQWRCRLRCLRQHQHDDFADDISLLK 456  
497 CALFMLPLCLMWQWRCRLRCLRQHQHDDFADDISLLK 532

RESULT 2  
Q9BYC1 PRELIMINARY; PRT; 476 AA.  
AC Q9BYC1; PRELIMINARY; PRT; 476 AA.  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE BETA-SITE APP CLEAVING ENZYME I-476.  
DE BACE.  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRIN;  
RX MEDLINE=21408467; PubMed=11516562;  
RA Tanahashi H., Tabira T.;  
RT "Three novel alternatively spliced isoforms of the human beta-site APP cleaving enzyme (BACE) and their effect on amyloid beta-peptide production.";  
RL Neurosci. Lett. 307:9-12(2001).  
DR EMBL; AB050436; BAB40931.1; -.  
DR HSSP; P32329; 1YPS.  
DR InterPro; IPR001969; Asp\_protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF00026; asp; 1.  
DR PRINTS; PR00792; PEPsin.  
DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
SQ SEQUENCE 476 AA; 52907 MW; 6C8C87F8A953AF66 CRC64;

Query Match 93.6%; Score 2264.5; DB 4; Length 476;  
Best Local Similarity 94.5%; Pred. No. 9.5e-185;  
Matches 431; Conservative 0; Mismatches 0; Indels 25; Gaps 1;  
QY 1 ETDEEPEEPCRRGSGFVEMVDNLRKSGQYVYVEMTVGSPQTLNILDVTGSSNFAVGAAP 60  
46 ETDEEPEEPCRRGSGFVEMVDNLRKSGQYVYVEMTVGSPQTLNILDVTGSSNFAVGAAP 105  
QY 61 HPFLHRYQRLSSTYRDLRKGVVVPTQCKWEGELGTLVSIPIHGPNTVTRANIAAITE 120



```
Db 46 ETDEEPEPCRRGGSFVEMVDNLKSGSQGYVYVMTGSPQTLNLLVDGSSNFVGAAP 105
QY 61 HPFLHRYQRLSSYTRDLRGKYVYPYTOGKWEGLGTDLVSIPHGPNVTVRANIAITE 120
Db 106 HPFLHRYQRLSSYTRDLRGKYVYPYTOGKWEGLGTDL----- 145
QY 121 SOKFFINGSNWEGILGLAYAEIARPDSDLPEFFDSLVKQTHVFNLSLQLCGAGPPLNQS 180
Db 146 -----PDSLEPPFDSLVKQTHVFNLSLQLCGAGPPLNQS 181
QY 181 EVLASVGSMIIGGIDHSLYTSGLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDK 240
Db 182 EVLASVGSMIIGGIDHSLYTSGLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDK 241
QY 241 SIVDSGTTNLRPLPKVFEAAVKSIIKAASSTKFPDGFGLGEOQVWQAGTTPWNIFPVIS 300
Db 242 SIVDSGTTNLRPLPKVFEAAVKSIIKAASSTKFPDGFGLGEOQVWQAGTTPWNIFPVIS 301
QY 301 LYLMEVNTNQSFRITILPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 360
Db 302 LYLMEVNTNQSFRITILPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 361
QY 361 FDRARRKIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDSTLMTIAYVMAAI 420
Db 362 FDRARRKIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDSTLMTIAYVMAAI 421
QY 421 CALFMLPLCLMVCQWRCRLRQHQHDDFADDISLLK 456
Db 422 CALFMLPLCLMVCQWRCRLRQHQHDDFADDISLLK 457

RESULT 4
Q9BYB9 ID Q9BYB9 PRELIMINARY; PRT; 432 AA.
AC Q9BYB9
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BETA-SITE APP CLEAVING ENZYME I-432.
GN BACE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=21408467; Pubmed=11516562;
RA Tanahashi H., Tabira T.;
RT "Three novel alternatively spliced isoforms of the human beta-site APP
RT cleaving enzyme (BACE) and their effect on amyloid beta-peptide
RT production.";
RL Neurosci. Lett. 307:9-12(2001).
RL EMBL; AB050438; BAB40933.1; -.
DR HSSP; P32329; 1YPS.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSTN.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN1.
SQ SEQUENCE 432 AA; 48212 MW; 96FC81B6F0EED01B CRC64;

Query Match 83.4%; Score 2016.5; DB 4; Length 432;
Best Local Similarity 84.9%; Pred. No. 1.2e-163;
Matches 387; Conservative 0; Mismatches 0; Indels 69; Gaps 1;

QY 1 ETDEEPEPCRRGGSFVEMVDNLKSGSQGYVYVMTGSPQTLNLLVDGSSNFVGAAP 60
Db 46 ETDEEPEPCRRGGSFVEMVDNLKSGSQGYVYVMTGSPQTLNLLVDGSSNFVGAAP 105
QY 61 HPFLHRYQRLSSYTRDLRGKYVYPYTOGKWEGLGTDLVSIPHGPNVTVRANIAITE 120
```

```
Db 106 HPFLHRYQRLSSYTRDLRGKYVYPYTOGKWEGLGTDL----- 145
QY 121 SDKFFINGSNWEGILGLAYAEIARPDSDLPEFFDSLVKQTHVFNLSLQLCGAGPPLNQS 180
Db 146 -----LCGAGPPLNQS 156
QY 181 EVLASVGSMIIGGIDHSLYTSGLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDK 240
Db 157 EVLASVGSMIIGGIDHSLYTSGLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDK 216
QY 241 SIVDSGTTNLRPLPKVFEAAVKSIIKAASSTKFPDGFGLGEOQVWQAGTTPWNIFPVIS 300
Db 217 SIVDSGTTNLRPLPKVFEAAVKSIIKAASSTKFPDGFGLGEOQVWQAGTTPWNIFPVIS 276
QY 301 LYLMEVNTNQSFRITILPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 360
Db 277 LYLMEVNTNQSFRITILPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 336
QY 361 FDRARRKIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDSTLMTIAYVMAAI 420
Db 337 FDRARRKIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDSTLMTIAYVMAAI 396
QY 421 CALFMLPLCLMVCQWRCRLRQHQHDDFADDISLLK 456
Db 397 CALFMLPLCLMVCQWRCRLRQHQHDDFADDISLLK 432

RESULT 5
Q9CUU5 ID Q9CUU5 PRELIMINARY; PRT; 266 AA.
AC Q9CUU5
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BETA-SITE APP CLEAVING ENZYME (FRAGMENT).
GN BACE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=BRAIN;
RX MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilmink L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK014390; BAB29317.1; -.
DR MEROPS; A01.004; -.
DR MGD; MGI:1346542; Bace.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 2.
FT NON_TER 1
SQ SEQUENCE 266 AA; 30201 MW; B913FDA8ADAB4238 CRC64;
```

Query Match		58.2%;	Score 1407;	DB 11;	Length 266;	
Best Local Similarity		98.9%;	Pred. No. 5.9e-112;			
Matches		263;	Conservative	1;	Mismatches	2; Indels 0; Gaps 0;
QY	191	IIIGDHSLSYTGSLWYTPIRREWYEVIIIVRVEINGDLMCKEYNDKSIYDSGTTNL	250			
Db	1	IIIGDHSLSYTGSLWYTPIRREWYEVIIIVRVEINGDLMCKEYNDKSIYDSGTTNL	60			
QY	251	RLPKVFEAAVKSIIKAASSTKFPDGFQWLGEOQLVCMWAGTTPWNIPFVLSLYLMGEVYNQ	310			
Db	61	RLPKVFEAAVKSIIKAASSTKFPDGFQWLGEOQLVCMWAGTTPWNIPFVLSLYLMGEVYNQ	120			
QY	311	SFRITILPQOYLREVDVATSDQCYKFAISQSTGTGVMGAVMEGFYVVDRAKRIGF	370			
Db	121	SFRITILPQOYLREVDVATSDQCYKFAISQSTGTGVMGAVMEGFYVVDRAKRIGF	180			
QY	371	AVSACHVDEFTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAIC-ALFMLPLCL	430			
Db	181	AVSACHVDEFTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAIC-ALFMLPLCL	240			
QY	431	MVCQWRLCLRLRQHQHDDFADDISLLK	456			
Db	241	MVCQWRLCLRLRQHQHDDFADDISLLK	266			
RESULT 6						
ID	Q9H2V8	PRELIMINARY;	PRT;	439 AA.		
AC	Q9H2V8;					
DT	01-MAR-2001 (TrEMBLrel. 16, Created)					
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)					
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)					
DE	CDAL3.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISUE=PHOECHROMOCYTOMA;					
RA	Li Y., Huang Q., Peng, Y., Song H., Yu Y., Xu S., Ren S., Chen Z.,					
RA	Han Z.;					
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; AF212252; AAG41783.1; -					
DR	HSSP; P00797; 2REN.					
DR	InterPro; IPR001969; Asp-protease.					
DR	InterPro; IPR001461; Pepsin.					
DR	Pfam; PF00026; asp; 3.					
DR	PRINTS; PR00792; PEPSIN.					
DR	PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.					
SQ	SEQUENCE 439 AA; 48275 MW; 02EC0E0E50F11602 CRC64;					
Query Match		48.0%;	Score 1160.5;	DB 4;	Length 439;	
Best Local Similarity		50.1%;	Pred. No. 1.3e-90;			
Matches		220;	Conservative	78;	Mismatches	134; Indels 7; Gaps 4;
QY	18	MYDNLRGKGGQGYVEMTVGSPPTINILVDTGSSNFAVGAAPHPFLHRYQQLSSYR	77			
Db	1	MYDNLGDSGRGYLEMLIGTPQKQLIQLVDTGSSNFAVGAAPHPFLHRYQQLSSYR	60			
QY	78	DLRKGVPYPTQKWEGLGTLVSIPIHGPNTVVRANIAITSDKPFNGSNWEGILGL	137			
Db	61	SKGFDVTVKYGTSWTFYVGEDLVITPKGNTSFLVNIAITFESNFPLPGIKWNGILGL	120			
QY	138	AVAEIARPDLSLEPPFDSLVKQTHVFNPLFSLQLCGAGFPLNQSEVLASVGGSMIIGDIDH	197			
Db	121	AYATLAKPSSLETFDLSLVQANIPNVFSMQMCGAGLPVAGS---GTNGGSLVGGIEP	177			
QY	198	SLYGLSWTPPIRREWYEVIIIVRVEINGDLMCKEYNDKSIYDSGTTNLRLPKVVF	257			
Db	178	SLYKGIWYTPKEEWYQTEILKLEIGGSLQNLDCREYNADKAIYDVSSTLLRLPQKVF	237			
Query Match		47.4%;	Score 1146.5;	DB 11;	Length 514;	
Best Local Similarity		50.7%;	Pred. No. 2.6e-89;			
Matches		216;	Conservative	75;	Mismatches	130; Indels 5; Gaps 3;
QY	8	EPGR-RGSFVEMVDNLGRKSGQGYVEMTVGSPPTINILVDTGSSNFAVGAAPHPFLHR	66			
Db	65	EPVRATANFLAMVDNLQDSCRGYLEMLIGTPQKQIQLVDTGSSNFAVGAAPHPFLHR	124			
QY	67	YQQLSLSTYRDLRKGVPYPTQKWEGLGTLVSIPIHGPNTVVRANIAITSDKXFFI	126			
Db	125	YFDESSTYHSKGFVDVTVKYGTSWTFYVGEDLVITPKGNTSFLVNIAITFESNFPL	184			
QY	127	NGSNWEGILGLAYAEIARPDLSLEPPFDSLVKQTHVFNPLFSLQLCGAGFPLNQSEVLASV	186			
Db	185	PGIKWNGILGLAYAEIARPDLSLEPPFDSLVKQTHVFNPLFSLQLCGAGFPLNQSEVLASV	241			
QY	187	GGSMIIGDHSLSYTGSLWYTPIRREWYEVIIIVRVEINGDLMCKEYNDKSIYDSG	246			
Db	242	GGSLVIGIEFSLYKGIWYTPKEEWYQTEILKLEIGGSLQNLDCREYNADKAIYDVS	301			
QY	247	TNLRPKVFEAAVKSIIKAASSTKFPDGFQWLGEOQLVCMWAGTTPWNIPFVLSLYLMGE	306			
Db	302	TTLLRLPQKVFDAVVEAVARTSLIEFSDGFWGTGAQLACWTNSETPTWYFPPKISYLRDE	361			
Query Match		47.4%;	Score 1146.5;	DB 11;	Length 514;	
Best Local Similarity		50.7%;	Pred. No. 2.6e-89;			
Matches		216;	Conservative	75;	Mismatches	130; Indels 5; Gaps 3;
QY	8	EPGR-RGSFVEMVDNLGRKSGQGYVEMTVGSPPTINILVDTGSSNFAVGAAPHPFLHR	66			
Db	65	EPVRATANFLAMVDNLQDSCRGYLEMLIGTPQKQIQLVDTGSSNFAVGAAPHPFLHR	124			
QY	67	YQQLSLSTYRDLRKGVPYPTQKWEGLGTLVSIPIHGPNTVVRANIAITSDKXFFI	126			
Db	125	YFDESSTYHSKGFVDVTVKYGTSWTFYVGEDLVITPKGNTSFLVNIAITFESNFPL	184			
QY	127	NGSNWEGILGLAYAEIARPDLSLEPPFDSLVKQTHVFNPLFSLQLCGAGFPLNQSEVLASV	186			
Db	185	PGIKWNGILGLAYAEIARPDLSLEPPFDSLVKQTHVFNPLFSLQLCGAGFPLNQSEVLASV	241			
QY	187	GGSMIIGDHSLSYTGSLWYTPIRREWYEVIIIVRVEINGDLMCKEYNDKSIYDSG	246			
Db	242	GGSLVIGIEFSLYKGIWYTPKEEWYQTEILKLEIGGSLQNLDCREYNADKAIYDVS	301			
QY	247	TNLRPKVFEAAVKSIIKAASSTKFPDGFQWLGEOQLVCMWAGTTPWNIPFVLSLYLMGE	306			
Db	302	TTLLRLPQKVFDAVVEAVARTSLIEFSDGFWGTGAQLACWTNSETPTWYFPPKISYLRDE	361			



```
ID Q9P0D2 PRELIMINARY; PRT; 213 AA.
AC Q9P0D2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HSPC104 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CORD BLOOD;
RA Zhang Q.H., Ye M., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
RA Fan H.Y., Mao Y.F., Dai M., Huang O.H., Chen S.J., Chen Z.;
RT "Human partial CDS cloned from cd34+ stem cells.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF161367; AAF28927.1;
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 1.
DR NON_TER 1
FT SEQUENCE 213 AA; 24338 MW; EC9D3FA31CFA835C CRC64;
SQ SEQUENCE 213 AA; 24338 MW; EC9D3FA31CFA835C CRC64;

Query Match 29.5%; Score 712.5; DB 4; Length 213;
Best Local Similarity 83.5%; Pred. No. 7.8e-53;
Matches 137; Conservative 4; Mismatches 12; Indels 11; Gaps 1;

Qy 193 GGIDHSLYTGSGLWTPIRREHYEYIIVRVEINGQDLKMDCKEYNDKSIYDVSQTTNRL 252
Db 1 GGIDHSLYTGSGLWTPIRREHYEYIIVRVEINGQDLKMDCKEYNDKSIYDVSQTTNRL 60
Qy 253 PKKFEAAVKSIAKASTKEKPDGFWLGEOLVCWQAGTTPNNIPFVLSLYLMGEVTTNOSF 312
Db 61 PKKFEAAVKSIAKASTKEKPDGFWLGEOLVCWQAGTTPNNIPFVLSLYLMGEVTTNOSF 120
Qy 313 RITILPOQYLPRVEDVATSDQCKYFAISQSSTGTVMGAVIMEG 356
Db 121 RITILPOQYLPR-----WKMMPRKTTVTVCHLTVIHG 153

RESULT 11
Q9R1P7 PRELIMINARY; PRT; 255 AA.
ID Q9R1P7
AC Q9R1P7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ASPARTYL PROTEASE (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Acarino M., Fumagalli P., Taramelli R., Ottolenghi S.;
RT "Cloning of a gene from chromosome 21 Down Region encoding a potential transmembrane protease.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF051150; AAD45964.1;
DR MEROPS; A01.041;
DR InterPro; IPR001969; Asp.protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 2.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
KW Protease.
FT NON_TER 1
SQ SEQUENCE 255 AA; 28685 MW; 53DE317815996D63 CRC64;

Query Match 24.7%; Score 596.5; DB 11; Length 255;
Best Local Similarity 47.8%; Pred. No. 8e-43;
```

```
Matches 109; Conservative 44; Mismatches 74; Indels 1; Gaps 1;

Qy 205 WYTPIRREHYEYIIVRVEINGQDLKMDCKEYNDKSIYDVSQTTNRLPKKFEAAVKS 264
Db 1 WYTPIRREHYEYIIVRVEINGQDLKMDCKEYNDKSIYDVSQTTNRLPKKFEAAVKS 60
Qy 265 KRAASSTEKPDGFWLGEOLVCWQAGTTPNNIPFVLSLYLMGEVTTNOSFRITILPOQYL 324
Db 61 ARTSLPEFSDGFWTGAQLACNTSETPWAYFPKTSIYLRDENASFRITILPOQYL 120
Qy 325 VEDVATSDQCKYFAISQSSTGTVMGAVIMEGFWVVDKRRKRGFAVSACHVDFRTA 384
Db 121 MGAAGENY-ECYRFGISSSTNALVIGATVMEGFWVVDKRRKRGFAVSACHVDFRTA 384
Qy 385 AVEGPEVTLDMEDCGYNIPQDTESTLMTIAYVMAATCALFMPLCLMV 432
Db 180 EISGPFSTEDIASNCVPAQALNEPILWIVSVYALMSVCGAILLVILL 227

RESULT 12
Q9VKP6 PRELIMINARY; PRT; 391 AA.
ID Q9VKP6
AC Q9VKP6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CG17134 PROTEIN.
GN CG17134.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktarglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fessler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleeb J.M., Reese M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
```



Query Match	12.9%	Score 312.5	DB 6	Length 386
Best Local Similarity	28.0%	Pred. No. 2.5e-18		
Matches	113	Conservative	69	Mismatches 131; Indels 91; Gaps 20;
QY	8	EPG--RRGSFVEMVNLGRKSGQGYVEMTGSPQTLNILDVTGSSNFAVGAAPHPFL--	64	
DB	35	EPAVRQGPPELLKNYMDAQ---YGEIGIGTPPQCFTTVFVDTGSANLWFSIHCKLLDI	91	
QY	65	----HRYYQRLSSTYRDLRKGVY--VPYTGKKEGELGTDLVSIHPGN-----VTVR	112	
DB	92	ACWTHRKYNSDKSSTY--VKNGTFTDIHYGSGSLGVLSDQTVSPVPCNPSSSPGGVTQ	149	
QY	113	ANI--AAITESDKFFINGSNWEGILGLAYARIAPDDSLFFPFFSLVKQTHV--PNLFSLQ	169	
DB	150	QRTFGEAKQGGVYFI--AAKFGGILGMAYPRIS--VNNLVFVFNLMQOKLVDKNVFS--	204	
QY	170	LCGAGFPLNQSEVLASVGGSMIGGIDHSLTGSLVTPPIRREWYEVIIIVRVEINGDGL	229	
DB	205	-----FFLNR--DPKAQPGGELMLGTSKYRGSIMFHNVTROYWQHMDQLDV--GSSL	257	
QY	230	KMDCKEYNDYKSIYDSGTNTNLRPKKVFEEAAVKSIAASTTEKFPDGFGLVGVQWAG	289	
DB	258	TV-CK--GGCEAIVDTGTSLVGPVEEVRELQAI-----G	290	
QY	290	TPWNIPVLSILYLMGEVTVNSFRITILPQOYLR--PYEDVATSDODCKYFAISQSSTGT-	347	
DB	291	AVP-----LIQGEYMPICEKVSSLPQVTVKLGKDYAXSPED--YALKVSAQATTVC	340	
QY	348	-----VMGAVIMEGFVYVDFRARKRIGFAVSA	374	
DB	341	LSGFMGMDIPPPGFLWLIGDVFYGRYYTVFEDQNRVGLAEAA	384	
RESULT	15			
Q9DEC2	ID	Q9DEC2	PRELIMINARY;	PRT; 384 AA.
AC	Q9DEC2;	2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001	(TrEMBLrel. 16, Last sequence update)		
DT	01-DEC-2001	(TrEMBLrel. 19, Last annotation update)		
DE	PEPSINOGEN A.			
OS	Xenopus laevis	(African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;			
OC	Xenopodinae; Xenopus.			
OX	NCBI_Taxid=8355;			
RP	SEQUENCE FROM N.A.			
PA	Ikuzawa M., Inokuchi T., Kobayashi K., Yasumasu S.;			
RT	"Molecular Cloning of Pepsinogens in Adult Xenopus laevis and Bullfrog			
RT	Rana catesbeiana.";			
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE			
CC	EUKARYOTIC ASPARTYL PROTEASES FAMILY.			
DR	EWBL; AB045380; BAB20798.1; --			
DR	HSSP; P00790; 1FSN.			
DR	MEROPS; A01.001; --			
DR	InterPro; IPR001969; Asp_protease.			
DR	InterPro; IPR001461; Pepsin.			
DR	Pfam; PF00026; asp. 1.			
DR	PRINTS; PR00792; PEPSIN			
DR	PROSITE; PS00141; ASP_PROTEASE; 1.			
KW	Aspartyl protease; Hydrolase.			
SQ	SEQUENCE 384 AA; 41888 MW; 729DD2E7953D9072 CRC64;			
Query Match	12.7%	Score 308	DB 13	Length 384;
Best Local Similarity	27.5%	Pred. No. 6e-18;		
Matches	99	Conservative	60	Mismatches 133; Indels 68; Gaps 16
QY	30	YVEMTVGSPQTLNILDVTGSSNFAVGAAPHPFL-----HRYYQRLSSTYRDLRK	82	

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: August 7, 2002, 09:13:35 ; Search time 22.58 Seconds  
(without alignments)  
493.272 Million cell updates/sec

Title: US-09-724-571-43  
Perfect score: 2419  
Sequence: 1 ETDEPEEPGRGSGFVEMVD.....CLRLRQHQDDFADDISLLK 456

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2413	99.8	501	4	US-09-009-191-2
2	2320	95.9	774	4	US-09-009-191-4
3	1171.5	48.4	518	3	US-08-999-723-2
4	1171.5	48.4	518	4	US-09-434-427-2
5	1146.5	47.4	514	4	US-09-717-432-2
6	1146.5	47.4	514	4	US-09-912-484-2
7	299.5	12.4	396	1	US-08-208-007A-13
8	299.5	12.4	396	4	US-09-032-523-9
9	295.5	12.2	412	1	US-08-208-007A-12
10	285.5	12.2	412	4	US-08-974-691-4
11	279.5	11.6	458	6	5217891-15
12	273.5	11.3	409	1	US-09-640-305-6
13	273.5	11.3	409	1	US-08-360-673-6
14	273.5	11.3	427	2	US-08-846-021A-8
15	270	11.2	410	1	US-08-088-633-2
16	270	11.2	410	1	US-08-245-756-2
17	270	11.2	410	1	US-08-441-750-2
18	270	11.2	410	2	US-08-441-751-2
19	270	11.2	410	5	PCT-US92-02521-2
20	250	10.3	349	4	US-09-032-523-3
21	227	9.4	398	1	US-08-328-314-2
22	227	9.4	398	1	US-08-731-045-2
23	212	8.6	437	4	US-08-974-691-3
24	210	8.7	397	3	US-09-079-415-2
25	196.5	8.1	430	1	US-08-535-237-2
26	195.5	8.1	427	1	US-07-958-222A-2
27	193	8.0	330	3	US-08-115-753-1

28	193	8.0	419	3	US-08-115-753-2	Sequence 2, Appli
29	193	8.0	419	3	US-08-115-753-33	Sequence 33, Appli
30	187	7.7	420	4	US-09-008-271A-4	Sequence 4, Appli
31	187	7.7	420	4	US-08-974-691-8	Sequence 8, Appli
32	184.5	7.6	395	1	US-08-723-938-3	Sequence 3, Appli
33	184.5	7.6	395	2	US-09-080-538-3	Sequence 3, Appli
34	184	7.6	445	4	US-08-974-691-6	Sequence 6, Appli
35	184	7.6	445	4	US-08-974-691-2	Sequence 6, Appli
36	149	6.2	437	4	US-09-353-332-2	Sequence 2, Appli
37	129.5	5.4	140	3	US-09-211-631-13	Sequence 13, Appli
38	129.5	5.4	140	4	US-09-265-628-13	Sequence 13, Appli
39	129.5	5.4	140	4	US-09-001-141-11	Sequence 11, Appli
40	129.5	5.4	140	4	US-09-532-803-6	Sequence 6, Appli
41	129.5	5.4	140	4	US-09-653-403-14	Sequence 14, Appli
42	97	4.0	1030	4	US-09-091-117-2	Sequence 2, Appli
43	95.5	3.9	280	4	US-09-160-246-14	Sequence 14, Appli
44	92.5	3.8	1097	2	US-08-680-326-39	Sequence 39, Appli
45	88	3.6	746	2	US-08-838-219B-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-09-009-191-2  
; Sequence 2, Application US/09009191  
; Patent No. 6319689  
; GENERAL INFORMATION:  
; APPLICANT: POWELL, DAVID  
; APPLICANT: CHAPMAN, CONRAD  
; APPLICANT: MURPHY, KAY  
; APPLICANT: SMITH, TRUDI  
; TITLE OF INVENTION: ASP2  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: RATNER & PRESTIA  
; STREET: P.O. BOX 980  
; CITY: VALLEY FORGE  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19482  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/009,191  
; FILING DATE: 20-JAN-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: UK 9701684.4  
; FILING DATE: 28-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PRESTIA, PAUL F  
; REGISTRATION NUMBER: 23,031  
; REFERENCE/DOCKET NUMBER: GH-70368  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-407-0700  
; TELEFAX: 610-407-0701  
; TELEX: 846169  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 501 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-009-191-2

Query Match 99.8%; Score 2413; DB 4; Length 501;  
Best Local Similarity 99.8%; Pred. No. 3.3e-251;

Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDEEPEPGRGGSFVEMVDNLGRKSGQGYVEMTVGSPQTNLILVDGTGSSNFVGAAP 60  
DB 46 ETDEEPEPGRGGSFVEMVDNLGRKSGQGYVEMTVGSPQTNLILVDGTGSSNFVGAAP 105  
QY 61 HPLHRYQRLSSTYRDLRKGVYPTQGWEGELGTDLSIPHGPNTVVRANIAAITE 120  
DB 106 HPLHRYQRLSSTYRDLRKGVYPTQGWEGELGTDLSIPHGPNTVVRANIAAITE 165  
QY 121 SDRFFINGSNWEGILGAYAEIARPDSDLVSKQTHVFNLSQLCGAGFLNQS 180  
DB 166 SDRFFINGSNWEGILGAYAEIARPDSDLVSKQTHVFNLSQLCGAGFLNQS 225  
QY 181 EVLASVGGSMIGIDHSLYTGSWTPPIRREWYEVIIIVRVEINGODLKMCKEYNYDK 240  
DB 226 EVLASVGGSMIGIDHSLYTGSWTPPIRREWYEVIIIVRVEINGODLKMCKEYNYDK 285  
QY 241 SYVDSGTTNLRPKKVEAAVKSIAASSTKFPDGFVLGEQLVCWQAGTTPWNIFFVIS 300  
DB 286 SYVDSGTTNLRPKKVEAAVKSIAASSTKFPDGFVLGEQLVCWQAGTTPWNIFFVIS 345  
QY 301 LYLMEVNTGSFRITILPQOYLRPVEDVATSDQDCYKFAISQSTGTVMGAVIMEGFYV 360  
DB 346 LYLMEVNTGSFRITILPQOYLRPVEDVATSDQDCYKFAISQSTGTVMGAVIMEGFYV 405  
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVGGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420  
DB 406 FDRARKRIGFAVSACHVHDEFRTAAVGGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 465  
QY 421 CALFMLPCLMVCQWRLCRLRQHQHDDFADDISLLK 456  
DB 466 CALFMLPCLMVCQWRLCRLRQHQHDDFADDISLLK 501

RESULT 2

US-09-009-191-4  
; Sequence 4, Application US/09009191  
; Patent No. 6319689  
; GENERAL INFORMATION:  
; APPLICANT: POWELL, DAVID  
; APPLICANT: CHAPMAN, CONRAD  
; APPLICANT: MURPHY, KAY  
; APPLICANT: SMITH, TRUDI  
; TITLE OF INVENTION: ASP2  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: RATNER & PRESTIA  
; STREET: P.O. BOX 980  
; CITY: VALLEY FORGE  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19482  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/009,191  
; FILING DATE: 20-JAN-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: UK 9701684.4  
; FILING DATE: 28-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PRESTIA, PAUL F  
; REGISTRATION NUMBER: 23,031  
; REFERENCE/DOCKET NUMBER: GH-70368  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-407-0700  
; TELEFAX: 610-407-0701  
; TELEX: 846169

; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 774 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-009-191-4

Query Match 95.9%; Score 2320; DB 4; Length 774;  
Best Local Similarity 98.9%; Pred No. 7.1e-241;  
Matches 439; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 13 GSFVEMVDNLGRKSGQGYVEMTVGSPQTNLILVDGTGSSNFVGAAPHFLHRYQRL 72  
DB 1 GSFVEMVDNLGRKSGQGYVEMTVGSPQTNLILVDGTGSSNFVGAAPHFLHRYQRL 60  
QY 73 SSTYRDLRKGVYPTQGWEGELGTDLSIPHGPNTVVRANIAAITESDKFFINGSNWE 132  
DB 61 SSTYRDLRKGVYPTQGWEGELGTDLSIPHGPNTVVRANIAAITESDKFFINGSNWE 120  
QY 133 GILGLAYAEIARPDSDLVSKQTHVFNLSQLCGAGFLNQSSEVLASVGGSMII 192  
DB 121 GILGLAYAEIARPDSDLVSKQTHVFNLSQLCGAGFLNQSSEVLASVGGSMII 180  
QY 193 GGIDHSLYTGSWTPPIRREWYEVIIIVRVEINGODLKMCKEYNYDKSIVDSGTTNLR 252  
DB 181 GGIDHSLYTGSWTPPIRREWYEVIIIVRVEINGODLKMCKEYNYDKSIVDSGTTNLR 240  
QY 253 PKKVFEAAVKSIAASSTKFPDGFVLGEQLVCWQAGTTPWNIFFPVISLYLMGEVNTQS 312  
DB 241 PKKVFEAAVKSIAASSTKFPDGFVLGEQLVCWQAGTTPWNIFFPVISLYLMGEVNTQS 300  
QY 313 RITILPQOYLRPVEDVATSDQDCYKFAISQSTGTVMGAVIMEGFYVVDRAKRIGFAV 372  
DB 301 RITILPQOYLRPVEDVATSDQDCYKFAISQSTGTVMGAVIMEGFYVVDRAKRIGFAV 360  
QY 373 SACHVHDEFRTAAVGGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAICALFMLPCLM 432  
DB 361 SACHVHDEFRTAAVGGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAICALFMLPCLM 420  
QY 433 COMFCLRLRQHQHDDFADDISLLK 456  
DB 421 COMFCLRLRQHQHDDFADDISLLK 444

RESULT 3

US-08-999-723-2  
; Sequence 2, Application US/08999723A  
; Patent No. 6025180  
; GENERAL INFORMATION:  
; APPLICANT: Powell, David J.  
; APPLICANT: Southan, Christopher  
; APPLICANT: Chapman, Conrad G.  
; APPLICANT: Evans, Joanne R.  
; TITLE OF INVENTION: ASP1  
; FILE REFERENCE: GH/0262  
; CURRENT APPLICATION NUMBER: US/08/999,723A  
; CURRENT FILING DATE: 1997-10-06  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 518  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-08-999-723-2

Query Match 48.4%; Score 1171.5; DB 3; Length 518;  
Best Local Similarity 49.6%; Pred. No. 2.2e-117;  
Matches 222; Conservative 80; Mismatches 139; Indels 7; Gaps 4;



Db 362 NASRFRITILPOLYTOPMMGAGFNY-ECYRFGISSSTNALVIGATVMEGFYVVFDRQR 420  
QY 367 RIGFAVSACHVHDEFTAAVEGPFVTLDMEDCGYNIPOQDESLMTIAYVMAAICAFML 426  
Db 421 RVGFVSPCAEIGTIVSEISGPFSTEDIASNCVPAQALNEPILMTIVSVYALMSVCGAILL 480  
QY 427 PLCLMV 432  
Db 481 VLILL 486  
RESULT 6  
US-09-912-484-2  
; Sequence 2, Application US/09912484  
; Patent No. 6358725  
; GENERAL INFORMATION:  
; APPLICANT: Christie, Gary  
; APPLICANT: Li, Xiaotong  
; APPLICANT: Powell, David J.  
; APPLICANT: Zhu, Yuan  
; TITLE OF INVENTION: Mouse Aspartic Secretase-1 (mASp1)  
; FILE REFERENCE: GP-70663-D1  
; CURRENT APPLICATION NUMBER: US/09/912,484  
; CURRENT FILING DATE: 2001-07-25  
; PRIOR APPLICATION NUMBER: 60/166,974  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: 09/717,432  
; PRIOR FILING DATE: 2000-11-21  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 514  
; TYPE: PRT  
; ORGANISM: MUS MUSCULUS  
US-09-912-484-2

Query Match 47.4%; Score 1146.5; DB 4; Length 514;  
Best Local Similarity 50.7%; Pred. No. 1.1e-114;  
Matches 216; Conservative 75; Mismatches 130; Indels 5; Gaps 3;  
QY 8 BPGR-RGSFVEMVDNLKSGQGYVMTGSPQTLNVLVDGTSSNFAVGAAPHPLHR 66  
Db 65 EPVRATANFLAMVDNLQDSCRGYLEMLICTPPKQVQLVDTGSSNFAVAGAPHSYIDT 124  
QY 67 YQROLSTYRDLRGVYVPTQCKWEGELGTDLVSIPIHGPNTVRANIAAITESDKFFI 126  
Db 125 YFDESSTYHSGFDVTVKYTGQSWTGFVGEDLVIPKGFNSFLVNIATIFESNEFL 184  
QY 127 NGSNNEGILGLAYAEIARPDSDLEPFDLSLVKQTHVPLNLSLQCGAGFPNQSEVLASV 186  
Db 185 PGIKWNGILGLAYALAKPSSLETFDLSVLAQAQKIPDIFSMQMGAGLPVAGS---GTN 241  
QY 187 GSGMIIGDHLSTYGLSWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIYDSG 246  
Db 242 GGSVLGGTEPLSYKGIWYTPKEWYQIIEILKEIGGONLNDREYNADRAIVDSG 301  
QY 247 TTNLRPKKVFEEAAKSIKAASSTKFFPDGFWLQVLCVQAGTTPWNIFFVISIYLME 306  
Db 302 TLLRLPQKVFDAVVEAVARTSLIPEFSDGFWTGAQACWTNSETPWAFPKISYLRDE 361  
QY 307 VTNQSFRTILPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVMEGFYVVFDRAR 366  
Db 362 NASRFRITILPOLYTOPMMGAGFNY-ECYRFGISSSTNALVIGATVMEGFYVVFDRQR 420  
QY 367 RIGFAVSACHVHDEFTAAVEGPFVTLDMEDCGYNIPOQDESLMTIAYVMAAICAFML 426  
Db 421 RVGFVSPCAEIGTIVSEISGPFSTEDIASNCVPAQALNEPILMTIVSVYALMSVCGAILL 480  
QY 427 PLCLMV 432  
Db 481 VLILL 486

RESULT 7  
US-08-007A-13  
; Sequence 13, Application US/08208007A  
; Patent No. 5501969  
; GENERAL INFORMATION:  
; APPLICANT: HASTINGS, ET AL.  
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/208,007A  
; FILING DATE: March 8, 1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER: No. 5501969e  
; FILING DATE: No. 5501969e  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY D.  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 325800-95  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 396 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-08-208-007A-13

Query Match 12.4%; Score 299.5; DB 1; Length 396;  
Best Local Similarity 25.9%; Pred. No. 1.3e-23;  
Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;  
QY 3 DEEPFEPGRGSGFVEMVDNLKSGQGYVMTGSPQTLNVLVDGTSSNFAVGA--- 58  
Db 63 DQSAKEP-----LNYLD-----MEYFGTISGSPQNFVIFDTGSSNLWVPSVYCT 110  
QY 59 APHPFLHRYQROLSTYRDLRGVYVPTQCKWEGELGTDLVSIPIHGPNTVRANIAAI 118  
Db 111 SPACKTHSRFPQSSYISQPGQSFISQYGTGSLGIAGADQSV-EGLTIVGQGFGEV 169  
QY 119 TESDKFFINGNWEGLGLAYAEIARPDSDLEPFDLSLVKQTHVPLNLSLQCGAGFPLN 178  
Db 170 TEPQGTFFV-AEFGILGLGYPPLA--VGGVTPVFDNMAO---NLVDLPMSFVYSSN 222  
QY 179 QSEVLASVGGSMIIGDHLSTYGLSWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNY 238  
Db 223 PE--GGAGSELIGGYDHSFSGSLNWPVTKQAYWQIALDNIQVG--TVMFCSE--G 275  
QY 239 DKSIVSDGTTNLRPKKVFEEAAKSIKAASSTKFFPDGFWLQVLCVQAGTTPWNIIPPV 298  
Db 276 COAIVDTGSLITGSPDKIKQLQNAIGAAP-----VDGEYAVE-----CANLNVMPD 322  
QY 299 ISLYLMGEVTNQSFRTILPQOYLRPVEDVATSDDCYKFAISQSSTG----- 346  
Db 323 VTFTING-----VPYTLSPYAT--TLILDVFDGMQFC-----SSGFGQGLDIHPPAGP 366

QY 347 -TVMGAVIMEGFYVDFRARRKRGFA 371  
Db 367 LWILGDFVFIQFYSVDFRGNRVGLA 392

RESULT 8  
US-09-032-523-9  
; Sequence 9, Application US/09032523  
; Patent No. 6232454  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Guegler, Karl  
; APPLICANT: Baugh, Mariah  
; TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/032,523  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0479 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 396 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: Genbank  
; CLONE: 181994  
US-09-032-523-9

Query Match 12.4%; Score 299.5; DB 4; Length 396;  
Best Local Similarity 25.9%; Pred. No. 1.3e-23;  
Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;

QY 3 DEEPEEPGRGSGFVMDNLRKSGGQYVEMTVGSPQTLNLTVDFTGSSNFAVGA----58  
Db 63 DQSAKEP-----LNYLD-----MEYFGTISIGSPQNTVIFDGGSNLWVPSVYCT 110  
QY 59 APHPFLHRYQRLSSYRDLRGVYVYVYQGWKEGELGDLVSIPIHGPNTVVRANIAAI 118  
Db 111 SPACKTHSRFPQSSTYSYQPGQSFQYGTGSLSGIGADQSV-EGLTVVQQGGEV 169  
QY 119 TESDKFFINGNNEGILGLAYAEIARPDSDLEPFDSLKYQTHVPLNLSLQCGAGFPLN 178  
Db 170 TEPQQTVD-AEFDGILGLCYPSLA--VCGVTPVFONMAQ-----NLVDLPMEFSVTWSSN 222

QY 179 QSEVLASVGGSMIIGIDHSYTGSLWYTPIRREWYVEVIVRVEINGQDLKMDCKEYNY 238  
Db 223 PE---GGAGSELIFGGYDHSFSGSLNWVPYTKQAYWQIALDNIOVGG--TVMECSE--G 275  
QY 239 DKSIVDSGTTNLRPKKVFEEAAVKAASSTKPKPDGFWLGEQLVCMQAGTTPWNIPPV 298  
Db 276 CQAIIVDTGSLITGSPDKIKQLQNAIGAAP-----VDGEYAVE-----CANLNVMPD 322  
QY 299 ISLYLMGEVTNQSPRITILPQOYLRPVEDVATSDQDCYKFAISQSSSTG-----346  
Db 323 VTFTING-----VPYTLSPYAT--TLLDVFDGMOFC-----SSGFOGLDIHPAPG 366  
QY 347 -TVMGAVIMEGFYVDFRARRKRGFA 371  
Db 367 LWILGDFVFIQFYSVDFRGNRVGLA 392

RESULT 9  
US-08-208-007A-12  
; Sequence 12, Application US/08208007A  
; Patent No. 5501969  
; GENERAL INFORMATION:  
; APPLICANT: HASTINGS, ET AL.  
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/208,007A  
; FILING DATE: March 8, 1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: No. 5501969e  
; FILING DATE: No. 5501969e  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY D.  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 325800-95  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 412 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-08-208-007A-12

Query Match 12.2%; Score 295.5; DB 1; Length 412;  
Best Local Similarity 28.5%; Pred. No. 3.8e-23;  
Matches 105; Conservative 58; Mismatches 146; Indels 59; Gaps 15;  
QY 30 YVEMTVGSPQTLNLTVDGSSNFAVGAAPHPFL-----HRYQRLSSSTYRDLRGV 83  
Db 79 YYGEIGTIGTPQCTVWEDTGSSNLWVPSIHKCLDIACWIHHKYNDSKSTYVKNGTSP 138  
QY 84 YVPYVYQGWKEGELGDLVSIPI-----HGPNTVVRANIAAITESDKFFINGSNMEGI 134  
Db 139 DIHYGSSLSGLYSODTVSVPCQSASSASALGGVKVQVFGEATKQPGITFIAAKFDGI 198

```

QY 135 LGLAYAEIARPDPSLEPFDDSLVKQTHV-PNLFSLQCGAGFPLNQLSEVLASVGGSMIIG 193
      ||||| : : : |||:|:| | | | | : | | | :|
Db 199 LGWAYPRIS--VNNVLPVFDNLMOQKLVQDNIFSYL-----SRDPDAQPGGELMIG 248
      ||||| : : : |||:|:| | | | | : | | | :|
QY 194 GIDHSLYTGSLWTPTRREWYVEIIVRVEI-NGQDLKMDCKEYNDKSTVDSGTTNLR 252
      ||||| : : : |||:|:| | | | | : | | | :|
Db 249 GTDSKYKGSLSYLNVRKAYQWVHLDOVEVASGLTL---CKE--GCEAIVDTGSLMVG 303
      ||||| : : : |||:|:| | | | | : | | | :|
QY 253 PKVFEAAVKSIAASSTEFKPDGFWLGEOLV-CWOAGTTPWNIFPVVISLYLMGEVTNOS 311
      ||||| : : : |||:|:| | | | | : | | | :|
Db 304 PVDEVRELOKAIGAVPLIQ-----GEYMIPCEKVST-----LPAITLKLGG----KG 346
      ||||| : : : |||:|:| | | | | : | | | :|
QY 312 FRITILPQOYLRPVEDVATSDDCYKFAISQ-----SSTGTVMGAVIMEGFYVVFDRARK 366
      ||||| : : : |||:|:| | | | | : | | | :|
Db 347 YKLS--PEDYTLKVSQAGKTL--CLSGFMGMDIPPPSGPLWILGDVFIGRYTVFDRDNN 402
      ||||| : : : |||:|:| | | | | : | | | :|
QY 367 RIGFAVSA 374
      ||||| :
Db 403 RVGFAEAA 410
      ||||| :

RESULT 10
US-08-974-691-4
; Sequence 4, Application US/08974691
; Patent No. 6225103
; GENERAL INFORMATION:
; APPLICANT: Keolsch, Gerald
; APPLICANT: Lin, Xinli
; APPLICANT: Tang, Jordan
; TITLE OF INVENTION: Cloning and Characterization of Napsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center, 1201 W. Peachtree
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,691
; FILING DATE: 20-NOV-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,196
; FILING DATE: 20-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/046,126
; FILING DATE: 09-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMRF 166
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-873-8794
; TELEFAX: 404-873-8795
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
;
US-08-974-691-4

```

```

Query Match 12.2%; Score 295.5; DB 4; Length 412;
Best Local Similarity 28.5%; Pred. No. 3.8e-23;
Matches 105; Conservative 58; Mismatches 146; Indels 59; Gaps 15;

QY 30 YVEMTVGSPPTNLILVDTGSSNFAVGAAPHFL-----HRYQROLSSYRDLRKG 83
      ||||| : : : |||:|:| | | | | : | | | :|
Db 79 YGEIGIGTPQCFVTVFDTGSSNLWVPSIHCKLLDIACIWHHKYNSDKSSTYVKNGTSTF 138
      ||||| : : : |||:|:| | | | | : | | | :|
QY 84 YVPYTGKWEGLGDLVSLIP-----HGPNTVVRANIAAITEESDKFFINGSNWEGI 134
      ||||| : : : |||:|:| | | | | : | | | :|
Db 139 DIHYGSGLSGLSQDTVSPQCQSASSASALGVKVRQVCEATKQGITFLAAKFDGI 198
      ||||| : : : |||:|:| | | | | : | | | :|
QY 135 LGLAYAEIARPDPSLEPFDDSLVKQTHV-PNLFSLQCGAGFPLNQLSEVLASVGGSMIIG 193
      ||||| : : : |||:|:| | | | | : | | | :|
Db 199 LGWAYPRIS--VNNVLPVFDNLMOQKLVQDNIFSYL-----SRDPDAQPGGELMIG 248
      ||||| : : : |||:~|:~| | | | | : | | | :|
QY 194 GIDHSLYTGSLWTPTRREWYVEIIVRVEI-NGQDLKMDCKEYNDKSTVDSGTTNLR 252
      ||||| : : : |||:|:~| | | | | : | | | :|
Db 249 GTDSKYKGSLSYLNVRKAYQWVHLDOVEVASGLTL---CKE--GCEAIVDTGSLMVG 303
      ||||| : : : |||:~|:~| | | | | : | | | :|
QY 253 PKVFEAAVKSIAASSTEFKPDGFWLGEOLV-CWOAGTTPWNIFPVVISLYLMGEVTNOS 311
      ||||| : : : |||:~|:~| | | | | : | | | :|
Db 304 PVDEVRELOKAIGAVPLIQ-----GEYMIPCEKVST-----LPAITLKLGG----KG 346
      ||||| : : : |||:~|:~| | | | | : | | | :|
QY 312 FRITILPQOYLRPVEDVATSDDCYKFAISQ-----SSTGTVMGAVIMEGFYVVFDRARK 366
      ||||| : : : |||:~|:~| | | | | : | | | :|
Db 347 YKLS--PEDYTLKVSQAGKTL--CLSGFMGMDIPPPSGPLWILGDVFIGRYTVFDRDNN 402
      ||||| : : : |||:~|:~| | | | | : | | | :|
QY 367 RIGFAVSA 374
      ||||| :
Db 403 RVGFAEAA 410
      ||||| :

RESULT 11
5217891-15
; Patent No. 5217891
; APPLICANT: BRAKE, ANTHONY J.;VAN DEN BERG, JOHAN A.
; TITLE OF INVENTION: DNA CONSTRUCTS CONTAINING A KLUYVEROMYCES
; A FACTOR LEADER SEQUENCE FOR DIRECTING SECRETION OF HETEROLOGOUS
; POLYPEPTIDES
; NUMBER OF SEQUENCES: 23
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/507,398
; FILING DATE: 09-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 78,551
; FILING DATE: 28-JUL-1987
; SEQ ID NO:15:
; LENGTH: 458
5217891-15

Query Match 11.6%; Score 279.5; DB 6; Length 458;
Best Local Similarity 26.8%; Pred. No. 2.4e-21;
Matches 95; Conservative 67; Mismatches 128; Indels 65; Gaps 15;

QY 31 YVEMTVGSPPTNLILVDTGSSNFAVGAAPHFL-----HRYQROLSSYRDLRKG 83
      ||||| : : : |||:~|:~| | | | | : | | | :|
Db 152 YFGIYLTPPQEFFTVLFDTGSSDEWV---PSIYCKSNACKNHQRFQDKSSTFQNLGKDL 208
      ||||| : : : |||:~|:~| | | | | : | | | :|
QY 84 YVPYTGKWEGLGDLVSLIPGPNVTVRANIAAITEESDKFFINGSNWEGILGLAYAEIA 143
      ||||| : : : |||:~|:~| | | | | : | | | :|
Db 209 SIHYGTSMQGITGYDTVTVSNIQDTVGLSTGEPGDFV--TYAEFDGILGWAYPSLA 266
      ||||| : : : |||:~|:~| | | | | : | | | :|
QY 144 RPDDSLPEPFDDSLVKQTHV-PNLFSLQCGAGFPLNQLSEVLASVGGSMIIGIDHSLYTG 202
      ||||| : : : |||:~|:~| | | | | : | | | :|
Db 267 ---SEYSIPVFDNMNRHLVAQDLFSVYMDRNG---QESMLT-----LGAIDPSYVTG 313
      ||||| : : : |||:~|:~| | | | | : | | | :|
QY 203 SLWYTPTRREWYVEIIVRVEINGQDLKMD--CKEYNDKSTVDSGTTNLRPKKVFEEA 260
      ||||| : : : |||:~|:~| | | | | : | | | :|
Db 314 SLHWVPVTQQYQWQFTVDSVTISGVVACEGCG-----QALDGTGSLKVGCPSSDIINI 367
      ||||| : : : |||:~|:~| | | | | : | | | :|

```



Query Match	11.28;	Score 270;	DB 1;	Length 410;
Best Local Similarity	25.7%;	Pred. No. 2.1e-20;		

Query Match	11.3%;	Score 273.5;	DB 2;	Length 421;
Best Local Similarity	26.4%;	Pred. No. 9.4e-21;		
Matches 94:	Conservative	68;	Mismatches 129;	
			Indels	65;
			Gaps	15;



```
Matches 91; Conservative 51; Mismatches 162; Indels 50; Gaps 10;
QY 30 YVEMTVGSPQTLNLYDTGSSNFAVCA---APHPFLHRYQRLSSYRDRLRKGYYV 85
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 96 YFEVSLGTTPQSFVILDTGSSNLWPSKDCGSLACFLHAKYDHDSSTYKKNSSFEI 155
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 86 PYTQKWEGLGTDLVSPHGPNTVVRANIAITAESDKFFINGSNWEGILGLAYAEIARP 145
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 156 RYGSMEGYVSQDVLIQI--GDLTPKVDFAEATSEPCLAFAFGKFDGILGLAY----- 207
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 146 DDSLEPPFDLSLVKQTHVPNLF---SLQLCGA---GFFLNQSEVLASVGGSMIIGDHS 199
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 208 -----DSISVVKIYPPYIKALELDLDPKFAFYLGTDKDESDGLATFGGVDKSK 259
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 200 YTGSLWYTPIRREWYEVIIIVREINGQDLKMDCKEYNDKSIVDGTTNLRPLPKVFEA 259
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 260 YEGKITWLPVRRKAYWE-----VSFDGVLGSEVAELOKTGAADTGTSIALPGLAEI 314
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 260 AVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIPFPVISLYLMGEVTNQSFRI 319
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 315 LNAEIGATKG-----WSGOYAVDCDTRDS---LPDLTLTFAG-----YNFTITPY 356
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 320 QYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFVVFDRARKRIGFAVS 373
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 357 DYTLEVSGSCISAFTPMDFP-EPIGPLAIGDSFLRKYYSVYDLGKDAVGLAKS 409
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```

Search completed: August 7, 2002, 09:14:48  
Job time: 73 sec

